HAL
open science

Analysis and approximation of compressible viscoplastic models with general nonlinearity for granular flows

Duc Hoai Nguyen

To cite this version:

Duc Hoai Nguyen. Analysis and approximation of compressible viscoplastic models with general nonlinearity for granular flows. Mathematics [math]. Université Gustave Eiffel, 2020. English. NNT : . tel-03078670

HAL Id: tel-03078670

https://theses.hal.science/tel-03078670

Submitted on 16 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

University of Paris-Est

Doctoral Thesis

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

in the field of

Mathematics

Presented by **Duc NGUYEN-HOAI**
Defended on December 14, 2020

Analysis and approximation of compressible viscoplastic models with general nonlinearity for granular flows

Composition of the jury

Prof. Carsten Carstensen	Humboldt University of Berlin	Reviewer
Prof. Laurent Chupin	University of Clermont Auvergne	Reviewer
Prof. Anne Mangeney	Institut de Physique du Globe de Paris	Examiner
Prof. Robert Eymard	Gustave Eiffel University	Examiner
Prof. François Bouchut	CNRS & Gustave Eiffel University	Advisor

Acknowledgements

First and foremost, I would like to express my deepest and most sincere gratitude to my supervisor Prof. François Bouchut for his continuous support of my PhD study and research, particularly for his patience, motivation, enthusiasm, and encouragement. His guidance helped me immensely during the entirety of the research and the writing of this thesis.
Besides, I would like to thank Prof. Carsten Carstensen and Prof. Laurent Chupin for their acceptance to be the reviewers of this manuscript. It is a great honor for me that they evaluate my work with profound care and interest. I also want to express my gratitude to Prof. Anne Mangeney and Prof. Robert Eymard for showing interest in this thesis by accepting to be a part of the jury panel.
Throughout four years of PhD, the thesis does not contain only the Mathematic content. I would like to thank the members of the LAMA for their welcoming, listening, their advice, and in particular Christiane Lafargue, Carole Auguin and Audrey Patout for taking charge of all the administrative procedures. Also, I am pleased to express my gratitude to Sylvie Cach for taking the time to answer many of my questions and helping me with the defense process. Thanks to all my friends at the LAMA; namely Yushun, Giulia, Evgenii, Thu-Huong, Arafat, Elias, Ahmed... for all the pleasant moments and all mathematical and non-mathematical discussions that I was able to share with them. Particularly, I would like to express my gratitude towards Prof. Marco Cannone, thanks to him, I had a chance to study Navier-Stokes equations, Besov and Sobolev spaces, and to meet my supervisor François Bouchut, all made possible per his recommendation . I also would like to thank Magdalena Kobylanski, Dominique Malicet, Luc Deleaval, for providing me with an enormous help in my teaching duties.
I also would like to thank CEMRACS 2019, through this summer program I have met a lot of people in my research domain or in related domain such as: Xavier, Meissa, Sergey, Jiao , Benoit, Francois... and specially Khawla Msheik, my best friend in the academic world. I also want to thank Christian Arber, my new supervisor in TopSolid for spending time to read my manuscript.
My thanks and appreciations also go to my circle of friends in France who have always been supportive and kind to me. Thanks to Minh Vuong and Van Than, although having researched on different topics, your passionate support have always enlightened me. Thanks to The Hung so much, you helped me countless times during my difficult years. Last but not least, this work could not have been possible without your support and patience, Marie. Thus, I would like to express my most sincere gratitude to you. I dedicate my last part to declare my deepest love to my beloved ones in my native country, although I would have preferred to express it in my native language.
Trước hết, con xin gửi lời cảm ơn sâu sắc tới bố mẹ vì những hi sinh, vất cả và cổ vũ của hai người. Cảm ơn chị Hiền và anh Đức, hai người vừa là gia đình, vừa là tấm gương, tiền bối cũng như đồng nghiệp của em, nhờ những chia sẻ của anh chị, em đã có thêm rất nhiều kinh nghiệm trong những năm tháng học tập của bản thân. Cảm ơn Việt, Sao, Bích, Subbie Ngô, mọi người dã rất kiên nhẫn với mình trong những năm qua. Đức cũng xin cảm ơn tới gia đình cô Thúy, chú Tiến, cô Thuận, chú Hân, cậu Minh, những người đã không những tạo điều kiện rất tất cho Đức học tập và giảng dạy những năm tháng còn ở Hà Nội, mà còn mang lại cho Đức thêm một gia đình thứ hai, có thêm nhiều người em như Sơn, Nam, Việt, An, Sóc,...Em xin gửi lời cám ơn tới khoa Toán của trường Đại học Khoa học tự nhiên Hà Nội, đặc biệt thầy Lê Minh Hà, cô Trịnh Thúy Giang, hai người luôn sẵn lòng chào đón và đưa ra cho em những lời khuyên kể cả tới khi em đã tốt nghiệp. Cảm ơn Giáo sư Đinh Nho Hào và các anh chị nhóm Phương trình vi phân - Phương trình Toán Lý của Viện Hàn Lâm Khoa học Việt Nam đã luôn chào đón em mỗi khi em trở về. Lời cuối cùng, quan trọng nhất, em xin gửi đến thầy Lê Văn An vì nếu không có thầy, mọi thứ đã không bắt đầu.

Abstract

This thesis is motivated by a research program between the LAMA (Mathematics, Univ. Gustave Eiffel) and the Institut de Physique du Globe of Paris (Earth Sciences) on granular media and their mathematical description.
We consider here a continuous description: the material is described as a fluid with viscoplastic rheology, that allows us to model the transition between static (solid) states and mobile (liquid) states. Incompressible models have been used since the introduction of the so called $\mu(I)$ rheology (Jop et al. 2006). However such models do not represent accurately real flows, even in laboratory experiments. Recent studies indicate that volume variations, even if not significantly large, play a key role in the dynamics. Therefore compressible models have been recently considered (Barker et al. 2017). Although particular rheologies such as Bingham or Herschel-Bulkley models have been often considered in mathematical studies such as Malek et al. 2010, not much can be found on general nonlinearities in terms of the trace and the norm of the strain rate tensor. We consider here compressible models with general nonlinearities $\sigma \in \partial F(D)$ where σ is the stress, D is the strain rate and F is a convex viscoplastic potential. Under technical assumptions on F such as subquadratic growth and superlinearity we prove the existence of solutions to the associated variational problem. This is obtained in the viscous as well as in the inviscid cases. We establish Euler-Lagrange characterizations of these solutions. No regularity is assumed on F, thus yield stress rheologies are included. Numerical methods for viscoplastic laws have been classically used: augmented Lagrangian or regularization methods. However these methods were designed merely for Bingham or Herschel-Bulkley fluids, and moreover their cost is still too high for applications to real configurations. Here we consider an iterative but explicit method in the sense that there is no linear system to solve, inherited from the minimizing of total variation functionals used in imaging (Chambolle, Pock 2011). It is applicable to any kind of nonlinearity, and includes a kind of projection on some convex sets. We prove the convergence of the method discretized in space with finite elements. Numerical tests confirm the theoretical results.

Résumé

Cette thèse est motivée par un programme de recherche entre le LAMA (Mathématiques, Univ. Gustave Eiffel) et l'Institut de Physique du Globe de Paris (Sciences de la terre) sur les milieux granulaires et leur description mathématique.
Nous considérons ici une description continue: le matériau est décrit comme un fluide avec rhéologie viscoplastique, qui permet de modéliser la transition entre les états statiques (solides) et mobiles (liquides). Des modèles incompressibles ont été utilisés depuis l'introduction de la rhéologie $\mu(I)$ (Jop et al. 2006). Néanmoins de tels modèles ne représentent pas correctement les écoulements réels, même dans les expériences de laboratoire. Des études récentes indiquent que les variations de volume, même si elles ne sont pas significativement grandes, jouent un rôle prépondérant dans la dynamique. Des modèles compressibles ont alors été considérés récemment (Barker et al. 2017). Bien que des rhéologies particulières telles que Bingham ou Herschel-Bulkley aient été souvent considérées dans les études mathématiques telles que Malek et al. 2010, peu d'études concernent des nonlinéarités générales en terme de la trace et de la norme du tenseur de taux de déformation. Nous considérons ici des modèles compressibles avec nonlinéarité générale $\sigma \in \partial F(D)$ où σ est le tenseur des contraintes, D le taux de déformation, et F est un potentiel viscoplastique convexe. Sous des hypothèses techniques sur F telles qu'une croissance sous-quadratique et une sur-linéarité, nous prouvons l'existence de solutions au problème variationnel associé. Cela est obtenu dans le cas visqueux aussi bien que dans le cas non visqueux. Nous établissons des caractérisations d'Euler-Lagrange de ces solutions. Aucune régularité n'est supposée sur F, et donc les rhéologies avec seuil de plasticité sont incluses. Des méthodes numériques pour les problèmes viscoplastiques ont été classiquement considérées: lagrangien augmenté ou régularisation. Cependant ces méthodes ont été mises au point essentiellement pour les fluides de Bingham ou d'Herschel-Bulkley, et de plus leur coût est encore trop élevé pour les applications aux configurations réelles. Ici nous considérons une méthode itérative mais explicite, dans le sens où il n'y a pas de système linéaire à résoudre, héritée de la minimisation de functionnelles de type variation totale utilisée en imagerie (Chambolle, Pock 2011). Elle est applicable à n'importe quelle forme de nonlinéarité, et fait intervenir une sorte de projection sur des ensembles convexes. Nous prouvons la convergence de la méthode discrétisée en espace avec des éléments finis. Des tests numériques confirment les résultats théoriques.

Contents

1 Introduction **1**
1.1 Physical origin . . . 1
1.2 Mathematical description of viscoplastic materials . . . 4
1.3 Numerical scheme and time algorithm . . . 9
1.4 The viscoplastic problem . . . 10
1.5 Space discretisation . . . 11
1.6 Sketch of the results . . . 12

2 Convex analysis and approximation methods for nonlinear viscoplastic models **13**
2.1 Preliminaries . . . 13
2.2 The regularization approach . . . 18
2.3 Augmented Lagrangian method . . . 20
2.4 Projection approach . . . 21
2.5 Pure plastic models . . . 23
2.6 Examples . . . 26

3 Existence of the solution for the steady viscoplastic model with nonlinear rheology **37**
3.1 With viscosity . . . 37
3.2 Without viscosity . . . 46

4 Explicit primal-dual algorithm **57**
4.1 Continuous formulation . . . 58
4.2 Finite element approximation . . . 64

5 Numerical experiments **73**
5.1 1D Bingham model with Euler transport . . . 73
5.2 2D compressible Bingham model . . . 79
5.3 Compressible Euler equations with Bingham viscoplasticity . . . 92

6 A lubrication equation for a simplified model of shear-thinning fluid **97**
6.1 Introduction . . . 97
6.2 Mathematical model . . . 98
6.3 Lubrication equation . . . 104
6.4 Numerical illustrations . . . 107

Bibliography **117**

Chapter 1

Introduction

1.1 Physical origin

Constitutive rheological laws for granular materials

Classical Physics contain several very well-known equations called "constitutive equations": Hooke's law, Fick's law, Ohm's law... These laws do not illustrate directly fundamental laws of nature, but rather a relation between two physical quantities, that should hold for any steady or dynamical evolution. In fluid dynamics, a crucial constitutive relation characterizing a material is called "Rheology". It means merely relations between stress and strain rate, and eventually other quantities. The history of this issue traces back to the 17th century. Sir Isaac Newton observed that each fluid has its own resistance to deformation at a given rate. It usually corresponds to the informal conception of the "thickness" of a fluid. For example, when pouring out a can of honey and another one containing water, it can be observed that flow of water can easily deform to a greater extent than the flow of honey does; thus one has tendency to conclude that Honey is "thicker" than water. To debunk this misconception, the quantity "viscosity" was created. Newton claimed that under certain ideal conditions each fluid has its own constant viscosity, for example, the statement "honey has a higher viscosity than water" is considered to be more precise. Now, if the fluid is examined under the assumption that its viscosity remains constant then it can be said intuitively that the flow with the higher velocity is "stronger". Scientifically one concludes that with higher strain rate (deformation over time), the fluids possesses higher stress tensor (internal forces that neighbouring particles of a continuous material exert on each other). There are fluids such as water, air, alcohol, glycerol, and thin motor oil, etc.. that have a linear stress-strain rate relation. Such single-phase fluids made up of simple, small-weight molecules are called "Newtonian fluids". Mathematically, the rheological relation can be formulated for Newtonian fluids as $\sigma = \eta Du$, where σ denotes the stress tensor and Du is the strain rate. However most fluids in the real world are Non-Newtonian fluids. Roughly speaking this means that the fluid has a shear stress nonlinearly proportional to the deformation of the media. A quite general constitutive equation can then be formulated as $\sigma = \widetilde{F}(Du)$ where $\widetilde{F}$ is a nonlinear function.

Bingham fluid

Even in the world of Non-Newtonian fluids, there are certain classifications based on the non-linearity of the stress - strain rate relation, such as: Fluids that exhibit a logarithmic increase in shear stress with shear strain rate are called shear-thinning fluids or pseudo-plastic fluids (i.e. blood, paint, ketchup). Meanwhile for other types of fluids which are called shear-thickening fluids

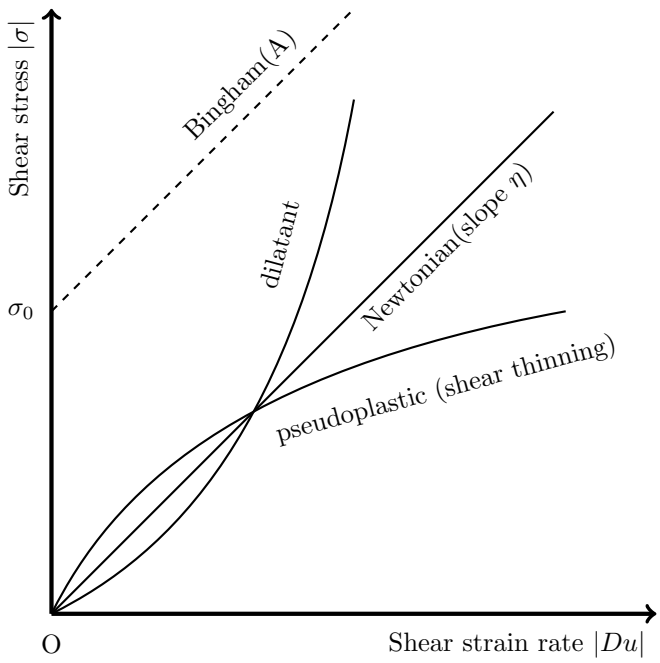


Figure 1.1: Time independent stress - strain rate relation for different types of fluids

(or dilatants), the viscosity increases exponentially with the increase in strain rate (i.e. honey, quicksand, oobleck, etc.). Various rheology types are sketched in Figure 1.1.
In 1913, [9, 8] a professor of the department of Chemistry at Lafayette College, named Eugene Cook Bingham introduced the first mathematical form describing the motion of a fluid that behaves as a rigid body at low stress but flows as a viscous fluid at high stress. This kind of Non-Newtonian fluid, also called "fluid by threshold", is named after him Bingham fluid, and constitutes a very important class of Non-Newtonian fluids. The study of Bingham materials is a crucial branch exhibiting the fundamental character of viscoplasticity. That is the reason why the simulation of Bingham fluids is always necessary and plays a key role in fully understanding the viscoplastic material; however, the work of modeling all the properties of viscoplastic media can not rely solely on the study of the Bingham model. Besides, modeling and simulating all the properties of viscoplastic media seems to be an impossible task. In general, besides the Non-Newtonian behaviour, this thesis will focus on the fact that viscoplastic fluids are also characterized by irreversible deformations inducing dissipation.

Why studying viscoplastic materials is important?

Viscoplastic media are present in our real life much more frequently than it is thought. An impressive example which proves that people cope with viscoplastic fluids everyday without even being aware of: Every morning, say you want to get some toothpaste out of the tube; when placed merely upside down, the paste will not come out even under the sole influence of the gravity or when squeezed with insufficient force. At this point, the toothpaste is considered to have an infinite viscosity and acts as a solid body. On the other hand, when the force is exerted beyond the certain threshold, the toothpaste will become runnier and will behave more "liquidy". There are numerous examples of viscoplastic media, namely butter, foams, pastes, slurries, oils, ceramics, etc. The study upon viscoplastic fluid has a vast number of applications in: the prediction of the plastic

Figure 1.2: Snow avalanche, pyroclast are also considered as viscoplastic media (Source: Internet)

collapse of structures, dynamic systems exposed to high strain rates, crash simulations, etc. We also can find their application in the petroleum industry: mitigation of paraffin wax, cements or drilling mud from crude oil. Moreover, a large variety of products in food industry (milk products, jam, chocolate confections, etc.) exhibit the fundamental characteristics of viscoplasticity. Their applications are also found in the study of environmental disasters (as in Figure 1.2): avalanche, ejection of volcanic magma, mudslides, and so forth.

This thesis is indeed motivated by the application to granular materials that are involved in the modeling of landslides. Their description by appropriate viscoplastic models is the subject of a research program between the Laboratoire d'Analyse et de Mathématiques Appliquées (Université Gustave Eiffel) and the Institut de Physique du Globe de Paris. A first relevant rheology for granular materials has been proposed by [37], the so called $\mu(I)$ rheology. This model does however not well represent real granular flows, even in experimental laboratory context. Additionally it has been proved by [3] that this model is strongly ill-posed. In numerical computations it can show shear band instabilities that are stronger and stronger when refining the mesh. The original $\mu(I)$ law being an incompressible model, it has been then recently shown [4] that considering compressible models could be a way to bypass the ill-posedness. Moreover volume variations, that do not exceed 10% in dry materials, are suspected to play nevertheless a key role in the dynamics. Moreover in the presence of an interstitial fluid like water (wet flows), volume and pore pressure variations are ubiquitous and their mathematical description are absolutely necessary. This is what is called "dilatancy", and a relevant law has been proposed in [51], that describes how the rheology should involve the divergence of the velocity field. Hence it is necessary to build models and develop simulations tools for viscoplastic compressible models with variable volume fraction. These models must include a yield stress to simulate the static and flowing parts of the material.

Progress in the mathematical study and numerical simulation of viscoplastic media

The rigorous analysis of viscoplastic models is at the moment mostly restricted to Bingham's law or direct generalizations. Over the past 40 years most studies assumed rheological models with a threshold of the stress tensor of Bingham or Herchel-Buckley type [34], where the viscosity follows

a power law, and depends only on the norm of the strain rate. Mathematical studies of compressible viscoplastic models can be found in [44, 52, 32, 45]. But most viscoplastic fluids are not like this, a dependency in the trace of the strain rate is important to include so called dilatancy effects, which are not present in most studies that consider incompressible models. The subject of this thesis is to discuss more complex models with nonlinearity that can include dilatancy effects. In particular the pure plastic models are part of it: Cam-Clay model, Drucker dilatant model or degenerate Bingham model. There is a theoretical issue in particular concerning the existence of solutions to these models and their characterization.

Numerically simulating viscoplastic flows is not straightforward. The difficulty is due to the presence of unknown interfaces separating the yielded and the unyielded regions, which are difficult to track. The problem can be written as a set of nonlinear variational inequalities. The case where it is well studied case is the Bingham case with viscosity [25, 29]. The most well-known method is the regularization method [57], but the augmented lagrangian method is also very much used [50]. These methods have not been extended to general nonlinear laws, and moreover the case without viscosity has not been much considered except [14].

1.2 Mathematical description of viscoplastic materials

1.2.1 Conservation of mass

Consider $\Omega \subset \mathbb{R}^N$ where $N = 1, 2, 3$ an open, bounded domain occupied by the medium. The conservation of mass expresses that the rate of increase of mass equals to the mass influx through the boundary of a control volume. Writing this law upon the control volume $\mathcal{V} \subset \Omega$ which is an arbitrary open, regular subdomain, this gives

$$\frac{d}{dt}\left(\int_{\mathcal{V}} \rho(t,x)\,\mathrm{dx}\right) = -\int_{\partial\mathcal{V}} \rho(t,x)u(t,x)\cdot\mathbf{n}(x)\,\mathrm{ds},$$

where $\mathbf{n}$ is the outward unit normal vector. The sign of the normal component of the velocity determines whether the fluid flows goes in or out of the control volume. Applying the Stokes formula this yields the concise form

$$\int_{\mathcal{V}} \left(\frac{\partial\rho}{\partial t} + \mathrm{div}(\rho u)\right)\mathrm{dx} = 0.$$

Since this relation can be applied for all arbitrary neighborhood of a point $x \in \Omega$ and for any $t \in (0,T)$, we obtain the local form of the conservation of mass in the differential form (known as the continuity equation)

$$\boxed{\frac{\partial\rho}{\partial t} + \mathrm{div}(\rho u) = 0 \quad \text{in } (0,T)\times\Omega \quad ,} \tag{1.2.1}$$

where $\rho = \rho(t,x)$ is the density of the flow, and $u = u(t,x) = (u_i(t,x))_{1\le i\le N}$ is the velocity vector field at point x and time t.

1.2.2 Conservation of momentum

The second relation is based on one of the most well-known law in physics: Newton's third law: the rate of increase of momentum equals to the net influx of momentum plus the exerted force. By

recalling the unit momentum per volume as ρu, we have the momentum in any control volume $\mathcal{V}$ would be $\int_{\mathcal{V}} \rho u \, \mathrm{dx}$.
The net inflow momentum can be written according to Stokes' formula

$$-\int_{\partial\mathcal{V}} \rho u (u \cdot \mathbf{n}) \, \mathrm{ds} = \int_{\mathcal{V}} \mathrm{div}(\rho u \otimes u) \, \mathrm{dx} \, .$$

The forces exerted upon the control volume $\mathcal{V}$ can be classified in two types of forces

- External force, in particular the total body force due to gravity. It can be formulated as

$$\int_{\mathcal{V}} \rho \, \mathrm{g},$$

  where g is the gravity vector that has constant magnitude.

- Internal force, caused by the deformation of the fluid, that can be written as

$$\int_{\partial\mathcal{V}} \sigma_{tot} \, \mathbf{n}(x) \, \mathrm{ds},$$

  where $\sigma_{tot}(t, x)$ is the symmetric tensor of total stress.

Finally we obtain the following formula which expresses the conservation of momentum

$$\int_{\mathcal{V}} \left(\frac{\partial(\rho u)}{\partial t} + \mathrm{div}(\rho u \otimes u) - \mathrm{div}\, \sigma_{tot} \right) \mathrm{dx} = \int_{\mathcal{V}} \rho \, \mathrm{g} \, \mathrm{dx} \, .$$

The local form of the conservation of momentum is therefore

$$\boxed{\frac{\partial(\rho u)}{\partial t} + \mathrm{div}(\rho u \otimes u) - \mathrm{div}\, \sigma_{tot} = \rho \, \mathrm{g} \quad \text{in } (0, T) \times \Omega.} \tag{1.2.2}$$

We shall assume from now on that the total stress can be decomposed as the sum of a compressible pressure law and a material dependent stress:

$$\boxed{\sigma_{tot} = -p_{th}(\rho) \, \mathrm{Id} + \sigma.} \tag{1.2.3}$$

1.2.3 Constitutive rheological law

The aim of this subsection is to discuss the heart of our subject: the law for σ. Let us start by some familiar notions in matrix theory. The **trace** of a matrix is defined as

$$\mathrm{Tr}(\sigma) := \sum_{i=1}^{N} \sigma_{ii}.$$

Let us denote by $\mathrm{Id} := (\delta_{ij})_{1 \le i,j \le N}$ the identity matrix, where δ_{ij} is the Kronecker symbol. Then we can always decompose a (symmetric) tensor σ as

$$\sigma = \frac{1}{N} \mathrm{Tr}(\sigma) \, \mathrm{Id} + \sigma', \tag{1.2.4}$$

where σ' (the "deviatoric" part of σ) satisfies $\mathrm{Tr}(\sigma') = 0$. We then introduce the pressure p by

$$p = -\frac{1}{N} \mathrm{Tr}(\sigma). \tag{1.2.5}$$

Combining (1.2.4) and (1.2.5) we get

$$\boxed{\sigma = -p\,\mathrm{Id} + \sigma', \qquad \mathrm{Tr}(\sigma') = 0.} \tag{1.2.6}$$

Let us remark that there is a general difficulty in defining the notion of pressure in fluids mechanics. It is surprising that such a classical notion remains not fully clear. Yet there is no concrete argument for formulating the pressure p as (1.2.5). One may wonder: Why do we have to decompose the diagonal part of the stress tensor? At least we can remark that if we define the inner product in the space of the tensors as

$$\sigma : \tau = \sum_{i,j=1}^{N} \sigma_{ij}\tau_{ij},$$

then it follows that the decomposition (1.2.4) is orthogonal. The associated Fröbenius norm $|\sigma| = (\sigma : \sigma)^{1/2}$ has the property to be a physical invariant: it can be computed with the same formula in any orthonormal basis.

The constitutive law expresses the relationship between the stress tensor σ and the **strain rate tensor**

$$Du = \frac{\nabla u + (\nabla u)^t}{2}. \tag{1.2.7}$$

Let us denote by $\mathbb{M}^s_{N\times N}(\mathbb{R})$ the space of real symmetric matrices of size $N \times N$. Suppose that $F : \mathbb{M}^s_{N\times N}(\mathbb{R}) \to \overline{\mathbb{R}}$ is convex, lower semi-continuous, and proper, where $\overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$. Let us recall that F is called **proper** if it is not identically $+\infty$ i.e. there exists at least one D such that $F(D) < +\infty$. We are going to consider rheologies of viscoplastic type defined by a relation

$$\boxed{\sigma \in \partial F(Du),} \tag{1.2.8}$$

where ∂F denotes the subdifferential, which is defined below. Knowing the monotonicity of the subdifferential of a convex function, (1.2.8) characterizes the monotonicity of the stress-strain rate relation. As we shall see this is an important property that ensures the existence of a variational formulation for the momentum evolution (1.2.2) (at least without inertial terms).

Since the above assumptions on F will be repeated many times, let us name them as

Hypothesis 1. *$F : \mathbb{M}^s_{N\times N}(\mathbb{R}) \to \overline{\mathbb{R}}$ is a convex, proper and lower semi-continuous function.*

Due to the lack of regularity for F, the notion sub-gradient/subdifferential ∂F which generalizes the classical gradient/differential, plays a crucial role. We denote by $\mathrm{dom}(F)$ the set of Ds that satisfy $F(D) < \infty$.

Definition 1.2.1 (Subgradient)**.** *Suppose that F satisfies Hypothesis 1. A symmetric tensor σ is called a **subgradient** of F at D if*

$$F(\overline{D}) \geq F(D) + \sigma : (\overline{D} - D), \qquad \forall \overline{D} \in \mathbb{M}^s_{N\times N}(\mathbb{R}).$$

Definition 1.2.2 (Subdifferential)**.** *Suppose that F satisfies Hypothesis 1. The **subdifferential** $\partial F(D)$ of F at D is the set of all subgradients:*

$$\partial F(D) = \Big\{\sigma \in \mathbb{M}^s_{N\times N}(\mathbb{R}) \mid F(\overline{D}) \geq F(D) + \sigma : (\overline{D} - D) \quad \forall \overline{D} \in \mathbb{M}^s_{N\times N}(\mathbb{R})\Big\}.$$

Taking into account this definition in the constitutive relation (1.2.8), we see that we obtain a set of inequalities, a so called variational formulation (here in a local setting, i.e. not integrated in space).

Let us introduce some well-known viscoplastic models that enter the considered class, i.e. they can be written as (1.2.8).

Examples

- Bingham model
  It takes the form (1.2.8) with

$$F(D) = \eta \frac{|D|^2}{2} + \sigma_0 |D| \qquad \forall D \in \mathbb{M}^s_{N\times N}(\mathbb{R}), \tag{1.2.9}$$

  where $\eta := \eta(\rho) > 0$ represents the viscosity of the fluid, and $\sigma_0 := \sigma_0(\rho) > 0$ is the yield stress. In other words the relation between σ and D is

$$\begin{cases} \sigma = \left(\eta + \frac{\sigma_0}{|D|}\right) D & \text{if } D \neq 0, \\ |\sigma| \leq \sigma_0 & \text{if } D = 0. \end{cases} \tag{1.2.10}$$

  Thus the material is solid ($D = 0$) for $|\sigma| \leq \sigma_0$, and liquid ($D \neq 0$) for $|\sigma| > \sigma_0$. The case without viscosity $\eta = 0$ is special: σ depends only on the direction of D, not on its magnitude. In this case it is a pure plastic model.

- Herschel-Bulkley model
  In this model F is taken as

$$F(D) = \frac{\eta}{1+d} |D|^{1+d} + \sigma_0 |D|,$$

  where $\eta > 0$, $d > 0$. Taking the subdifferential of F gives

$$\begin{cases} \sigma = (\eta |D|^{d-1} + \sigma_0 |D|^{-1}) D & \text{if } D \neq 0, \\ |\sigma| \leq \sigma_0 & \text{if } D = 0. \end{cases}$$

  Again σ_0 is the yield stress, η is the consistency parameter, $d > 0$ is the flow index. Roughly speaking, for $d < 1$ the fluid exhibits shear-thinning properties whereas for $d > 1$ it is shear-thickening. The case $d = 1$ corresponds to the Bingham model.

- Newtonian model
  The simplest case in the Herschel-Bulkley model is to take $n = 1$ and $\sigma_0 = 0$, i.e. $F(D) = \eta |D|^2/2$. The model then reduces to the well-known Newtonian model $\sigma = \eta D$ (Navier-Stokes equations).

The previous models are rather simple: F depends only on $|D|$. There exist more elaborate models which are relevant for granular media: Cam-Clay model, Drucker-dilatant model, degenerate Bingham model. These will be discussed in Chapter 2. In general, the frame invariance principle says that F should depend only the quantities $\mathrm{Tr}(D^k)$ for $k = 1, \ldots, N$. If $N = 2$ this means that F can depend on $\mathrm{Tr}(D)$ and $|D|$, or equivalently on $\mathrm{Tr}(D)$ and $|D'|$ where D' is the deviatoric part of D because $|D|^2 = \frac{(\mathrm{Tr}(D))^2}{N} + |D'|^2$. If $N = 3$ there is an additional possible dependency in $\det D$.

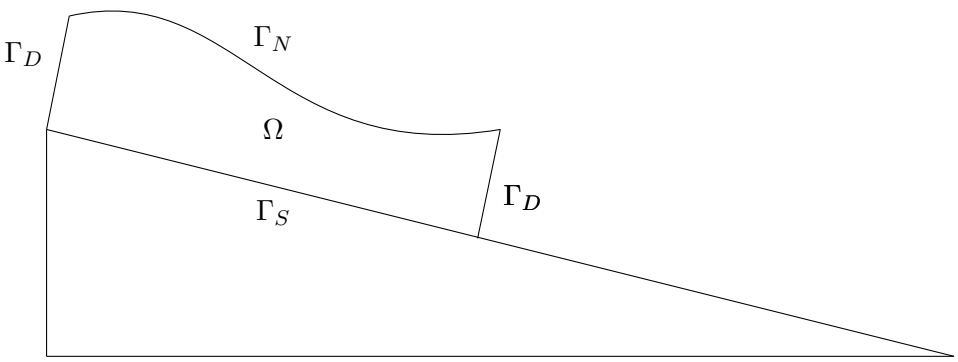


Figure 1.3: A domain with several boundary parts

For mechanically relevant models a difficulty is that the rheology is often formulated in different variables. Indeed decomposing σ and D by their trace and deviatoric parts

$$\sigma = -p\,\mathrm{Id} + \sigma', \qquad D = \frac{\mathrm{Tr}(D)}{N}\,\mathrm{Id} + D', \tag{1.2.11}$$

the relation between σ and D when F depends only on $\mathrm{Tr}(D)$ and $|D'|$ can be written as σ' and D' are colinear with same direction and (in the univalued part of ∂F)

$$p = f_1(\mathrm{Tr}(D), |D'|), \quad |\sigma'| = f_2(\mathrm{Tr}(D), |D'|), \tag{1.2.12}$$

with f_1, f_2 two scalar functions. The mechanical description of rheology is more often written as

$$|\sigma'| = g_1(p, |D'|), \quad \mathrm{Tr}(D) = g_2(p, |D'|), \tag{1.2.13}$$

with g_1, g_2 two scalar functions that represent respectively the internal friction law and the dilatancy law. It is not so easy to pass from (1.2.13) to (1.2.12).

1.2.4 Initial and boundary conditions

The problem (1.2.1), (1.2.2) is completed by initial conditions on the density and the velocity field,

$$\begin{aligned} \rho(t=0,x) &= \rho_0(x), \quad x \in \Omega, \\ u(t=0,x) &= u_0(x), \quad x \in \Omega. \end{aligned} \tag{1.2.14}$$

One has also to set conditions on the boundary of Ω. There are several boundary conditions that are physically relevant: Dirichlet, Neumann, periodic, slip, friction conditions. To give an example, let us consider that the boundary is decomposed as several parts and impose a different condition on each of them, see Figure 1.3. It can be

$$\begin{array}{lrl} \textbf{Dirichlet} & u = 0 & \text{in } (0,T) \times \Gamma_D, \\ \textbf{Neumann} & \sigma n = 0 & \text{in } (0,T) \times \Gamma_N, \\ \textbf{slip} & u \cdot n = 0 \text{ and } (\sigma n) \times n = 0 & \text{in } (0,T) \times \Gamma_S, \end{array} \tag{1.2.15}$$

where $\partial\Omega = \Gamma_D \cup \Gamma_N \cup \Gamma_S$. For applications to granular flows it is useful to have a moving free boundary. In this thesis we shall not consider this case.

1.2.5 Incompressible and compressible models

One states that "The compressibility (change in volume due to change in pressure) of a liquid is inversely proportional to its volumic modulus of elasticity, also known as the bulk modulus". In simple terms, the compressibility of the fluid is related to the variation of its density ρ. When the density is constant, the continuity equation (1.2.1) reduces to

$$\operatorname{div} u = 0. \tag{1.2.16}$$

Imposing this condition, the strain rate $D = Du$ is no longer an arbitrary symmetric matrix, but it has vanishing trace: $\mathrm{Tr}(D) = 0$. Therefore in this case the constitutive relation (1.2.8) needs to be understood with the subdifferential holding in the vector space of symmetric trace free matrices. It follows that the pressure p is not determined by the constitutive relation. Instead it appears as a Lagrange multiplier for the constraint (1.2.16). Thus the incompressible formulation is slightly different from the compressible formulation described above. In this thesis we consider only the compressible formulation.

1.2.6 Problem statement and state of the art

Combining (1.2.1), (1.2.2), (1.2.3), (1.2.8), (1.2.14), (1.2.15), our viscoplastic problem writes: Find (ρ, u) satisfying

$$\begin{cases} \dfrac{\partial \rho}{\partial t} + \operatorname{div}(\rho u) = 0 & \text{in } (0,T)\times\Omega, \\ \dfrac{\partial(\rho u)}{\partial t} + \operatorname{div}(\rho u \otimes u) + \nabla p_{th}(\rho) - \operatorname{div}\sigma = \rho g & \text{in } (0,T)\times\Omega, \\ \sigma \in \partial F(Du) & \text{in } (0,T)\times\Omega, \\ u = 0 & \text{in } (0,T)\times\Gamma_D, \\ \sigma n = 0 & \text{in } (0,T)\times\Gamma_N, \\ u\cdot n = 0 \text{ and } (\sigma n)\times n = 0 & \text{in } (0,T)\times\Gamma_S, \\ \rho(t=0,x) = \rho_0(x),\ u(t=0,x) = u_0(x), & x\in\Omega. \end{cases} \tag{1.2.17}$$

where F satisfies Hypothesis 1. To simplify we can consider that $p_{th}(\rho) = \kappa\rho^\gamma$ where $\gamma > 1$, a classical behaviour in gas dynamics. Physically relevant values for gases are $\gamma = \dfrac{5}{3}$ or $\dfrac{7}{5}$.
The previous system of equations is classical in fluid mechanics, except the rheological behaviour. In the case of a Newtonian material $\sigma = \eta Du$, this is the compressible Euler system of gas dynamics. Existence of solutions to this system in the Newtonian case is now known in arbitrary dimension for γ not too small. However this is still mostly an open problem for non Newtonian rheologies, see however [27, 45], and [5] for the 1d case. Notice that the existence for the 2d incompressible inviscid Bingham including inertial terms has been established in [39], see also [23].

1.3 Numerical scheme and time algorithm

In order to solve numerically the problem (1.2.17), we use the standard time splitting algorithm. Let N_t be a positive integer. We consider a constant timestep $\Delta t > 0$ and define the discrete times by

$$t_n = n\Delta t, \quad n \in \{0, 1, \dots, N_t\}.$$

We denote by u^n the approximation of the velocity field u at time t_n, i.e. $u^n \approx u(t_n)$, and similarly for the density $\rho^n \approx \rho(t_n)$. We use a similar notation for the stress tensor $\sigma^n \approx \sigma(t_n)$ or other useful quantities. We shall also need quantities $u^{n+\frac{1}{2}}$, $\rho^{n+\frac{1}{2}}$ that are intermediate between n and $n+1$.

In order to update the values ρ^n and u^n we proceed as follows. Suppose that ρ^n, u^n are known. As a first step we apply the finite volume method (FVM) to solve the compressible Euler system, and get $u^{n+\frac{1}{2}}, \rho^{n+\frac{1}{2}}$. In other words we have a finite volume discretization of

$$\begin{cases} \dfrac{\rho^{n+\frac{1}{2}} - \rho^n}{\Delta t} + \operatorname{div}(\rho^n u^n) = 0, \\[2ex] \dfrac{\rho^{n+\frac{1}{2}} u^{n+\frac{1}{2}} - \rho^n u^n}{\Delta t} + \operatorname{div}(\rho^n u^n \otimes u^n + p^n_{th}\,\mathrm{Id}) = 0. \end{cases} \tag{1.3.1}$$

In the second step we obtain u^{n+1}, ρ^{n+1} by using the finite element method (FEM) to solve

$$\begin{cases} \dfrac{\rho^{n+1} - \rho^{n+\frac{1}{2}}}{\Delta t} = 0, \\[2ex] \dfrac{\rho^{n+1} u^{n+1} - \rho^{n+\frac{1}{2}} u^{n+\frac{1}{2}}}{\Delta t} - \operatorname{div}\sigma^{n+1} = f^{n+\frac{1}{2}}, \quad \sigma^{n+1} \in \partial F(Du^{n+1}), \end{cases} \tag{1.3.2}$$

with $f^{n+1/2} = \rho^{n+1/2} g$. Since the first equation gives trivially $\rho^{n+1} = \rho^{n+1/2}$, only the second equation comes into play, it is a viscoplastic problem with a space dependent weight (the density). We also need boundary conditions for each of the systems (1.3.2) and (1.3.1). We will discuss this in more detail in Chapter 5.

1.4 The viscoplastic problem

In the previous subsection we have seen that the time splitting algorithm leads to the viscoplastic problem (1.3.2). When the weight $\rho^{n+1/2}$ is constant ($= 1$) this can be written as a steady problem

$$\alpha u - \operatorname{div}\sigma = f, \qquad \sigma \in \partial F(Du), \tag{1.4.1}$$

with $\alpha = 1/\Delta t$ and $f = f^{n+1/2} + \frac{u^{n+1/2}}{\Delta t}$. This viscoplastic problem can be formally obtained by minimizing the functional

$$J(v) = \alpha \int_\Omega \frac{|v|^2}{2} + \int_\Omega F(Dv) - \int_\Omega f \cdot v. \tag{1.4.2}$$

This minimization can be proved to be equivalent to the variational formulation

$$\alpha \int_\Omega u \cdot (v-u) + \int_\Omega F(Dv) \geq \int_\Omega F(Du) + \int_\Omega f \cdot (v-u), \quad \text{for all functions } v. \tag{1.4.3}$$

This problem has been studied a lot [25, 57, 58], but usually only the Bingham model with positive viscosity is considered, or eventually the Herschel-Bulkley model. The case of inviscid Bingham has been considered in [14], where the time dependent case is also treated. In this thesis we consider a quite general F, with or without viscosity. Numerical methods to solve (1.4.1) are numerous, and we refer to Chapter 2 for the description of classical ones.

1.5 Space discretisation

As mentioned above the general scheme uses a splitting algorithm in order to solve two simpler problems, one by the FVM and the other by the FEM. We use the approach of [12, 13] that deals with a nonlinear hyperbolic scalar conservation law regularised by the total variation flow operator. Here we consider a compressible fluid model together with viscoplastic rheology.
First we have to recall the space discretisation used in those articles, that allows us to create two dual meshes, one for the FVM and the other for the FEM.

The finite element mesh, denoted by $\mathcal{T}_h$, is a conforming mesh of the open, bounded domain $\Omega \subset \mathbb{R}^N$ of size h: $\mathcal{T}_h$ is a finite set of disjoint open simplices such that $\cup_{K\in\mathcal{T}_h}\overline{K} = \overline{\Omega}$, and h is the maximum value of the diameter of all $K \in \mathcal{T}_h$. The mesh is conforming in the sense that for two distinct elements K, L of $\mathcal{T}_h$, $K \cap L$ is either empty or a simplex included in an affine subset with dimension strictly lower than N, whose vertices are simultaneously vertices of K and L. The finite set of the vertices of the mesh is denoted by $\{x_i, i \in V\}$. Each element of $\mathcal{T}_h$ is assumed to be nonobtuse; which means that the angles between any two facets are less than or equal to $\pi/2$. For any $K \in \mathcal{T}_h$, we denote by $\mathcal{V}_K \subset V$ the set of the $N+1$ indices of the vertices of K, and by $\mathcal{E}_K$ the set of the $N+1$ faces of K.

The finite volume mesh, denoted by $\mathcal{D}_h$, is a polyhedral mesh of $\overline{\Omega}$ such that the interface between two cells is a finite union of faces. The mesh $\mathcal{D}_h$ is a dual mesh of $\mathcal{T}_h$ in the sense that each cell of $\mathcal{D}_h$ contains one and only one node of $\mathcal{T}_h$. For any $i \in V$, the cell of $\mathcal{D}_h$ containing the node x_i is denoted by Q_i. We assume that

$$\forall i \in V, Q_i \subset \bigcup_{K\in\mathcal{T}_h, i\in\mathcal{V}_K} \overline{K}. \tag{1.5.1}$$

Besides, N_i is the set containing the indices of the neighbouring cells of Q_i, $\mathcal{E}_h$ is the set of couples (i, j) such that Q_i and Q_j are neighbours and $i < j$, $\Gamma_{i,j}$ is the interface between two neighbour cells Q_i and Q_j, $\mathbf{n}_{i,j}$ is the unit normal vector to $\Gamma_{i,j}$ pointing toward Q_j.

The unknown function $u(t, x)$ is reconstructed simultaneously from the values $u_h = (u_i)_{i\in V}$ at the points x_i for $i \in V$, using a continuous piecewise affine reconstruction denoted by $\hat{u}_h$, and using a piecewise constant reconstruction denoted by $\bar{u}_h$. More explicitly, for all $u_h \in \mathbb{R}^V$ we have

$$\begin{aligned}
&\hat{u}_h \in C(\overline{\Omega}), \quad \hat{u}_h\,|_K \text{ is affine for each cell } K \in \mathcal{T}_h, \quad \hat{u}_h(x_i) = u_i \quad \forall i \in V,\\
&\bar{u}_h \in L^1_{loc}(\Omega), \quad \bar{u}_h(x) = u_i \ \forall x \in Q_i, i \in V.
\end{aligned}$$

Similarly, we use the same notation for $\hat{\rho}_h$, $\bar{\rho}_h$.

The numerical scheme approximating (1.2.17) is defined by
1 - Initialisation of $u_h^0 \in \mathbb{R}^V$, $\rho_h^0 \in \mathbb{R}^V$:

$$\begin{aligned}
u_i^0 &= u_{init}(x_i) \quad \forall i \in V,\\
\rho_i^0 &= \rho_{init}(x_i) \quad \forall i \in V.
\end{aligned}$$

2 - Finite volume step: Suppose that (u_h^n, ρ_h^n) are known. We compute the numerical flux by a first-order explicit formula

$$F_{i,j}^n = F(U_i^n, U_j^n, \mathbf{n}_{i,j}) \tag{1.5.2}$$

corresponding to each pair of neighbor cells Q_i, Q_j, where $U_i = (\rho_i, \rho_i u_i)$. We assume that the scheme is conservative, i.e. $F_{j,i}^n = -F_{i,j}^n$. We impose the CFL condition (for Courant, Friedrichs,

Levy [24]) on the timestep to prevent the blow up of the numerical values, under the form

$$\Delta t a_{ij}^n \leq h, \tag{1.5.3}$$

at each interface between two cells Q_i, Q_j, where a_{ij} is an approximation of the propagation speed, that should be deduced from the choice of the numerical flux. We compute $\rho_h^{n+\frac{1}{2}}$, $u_h^{n+\frac{1}{2}}$ by

$$U_i^{n+\frac{1}{2}} = U_i^n - \frac{\Delta t}{|Q_i|} \sum_{j \in N_i} |\Gamma_{i,j}| F_{i,j}^n. \tag{1.5.4}$$

3 - Finite element step: We set $\rho_h^{n+1} = \rho_h^{n+\frac{1}{2}}$. Then, define

$$\begin{aligned} V_h &:= \{\hat{v}_h \in C(\overline{\Omega}) \text{ such that } \hat{v}_h\,|_K \text{ is affine for each cell } K \in \mathcal{T}_h\}, \\ \Lambda_h &:= \{\sigma_h \in L^\infty(\Omega) \text{ such that } \sigma_h\,|_K \text{ is constant for each } K \in \mathcal{T}_h\}. \end{aligned}$$

The values $\rho_h^{n+\frac{1}{2}}$, $u_h^{n+\frac{1}{2}}$ being known, we look for $(\hat{u}_h^{n+1}, \sigma_h^{n+1}) \in V_h \times \Lambda_h$ such that

$$\int_\Omega \bar{\rho}_h^{n+1/2} \frac{\bar{u}_h^{n+1} - \bar{u}_h^{n+\frac{1}{2}}}{\Delta t} \cdot \bar{v}_h \,\mathrm{dx} + \int_\Omega \sigma_h^{n+1} : D\hat{v}_h \,\mathrm{dx} = \int_\Omega \bar{f}_h \cdot \bar{v}_h \,\mathrm{dx} \quad \forall \hat{v}_h \in V_h, \tag{1.5.5}$$

$$\sigma_h^{n+1} \in \partial F(D\hat{u}_h^{n+1}) \text{ a.e. in } \Omega. \tag{1.5.6}$$

1.6 Sketch of the results

Chapter 2 describes mathematical tools that are useful to study (1.4.1), and classical numerical methods to solve this problem. A particular attention is given to the description of pure plastic models. Then Chapter 3 concerns the rigorous analysis of the theoretical problem (1.4.1). Under technical assumptions on F such as subquadratic growth and superlinearity we prove the existence of solutions to the associated variational problem (1.4.3). This is obtained in the viscous as well as in the inviscid cases. We establish Euler-Lagrange characterizations of these solutions, i.e. the equivalence with the local formulation (1.4.1). No regularity is assumed on F, thus yield stress rheologies are included. Then Chapter 4 introduces an iterative but explicit method in the sense that there is no linear system to solve, inherited from the minimizing of total variation functionals used in imaging [20]. It is applicable to any type of nonlinearity, and includes a kind of projection on some convex sets, that already appears in the augmented Lagrangian method or in the time regularized approach of [22]. We prove the convergence of the method discretized in space with finite elements. Numerical tests performed in Chapter 5 confirm the theoretical results. Finally Chapter 6 is a work on a lubrication equation for a simplified model of shear-thinning fluid, that has been performed at the CEMRACS 2019 in collaboration with Khawla Msheik, Meissa M2019Baye and François James.

Chapter 2

Convex analysis and approximation methods for nonlinear viscoplastic models

In this section we consider the steady viscoplastic model

$$\begin{cases} \sigma \in \partial F(Du), & \\ \alpha u - \operatorname{div} \sigma = f & \text{in} \quad \Omega, \\ u = 0 & \text{on} \quad \Gamma, \end{cases} \tag{2.0.1}$$

where α is a positive constant, F satisfies Hypothesis 1 and f is a given source. We are going to introduce briefly the state-of-the-art concerning the numerical resolution of (2.0.1), and introduce useful tools of convex analysis.

There are various methods concerning the numerical simulation of viscoplastic flows, in particular for the Bingham fluid when $F(D) = |D|$. These results have been the subject of intense investigations from the early 1970 up to now. There are mainly two historical methods to approximate viscoplastic models. The first is based on the regularization procedure, and the second is based on the use of duality formulations, as the augmented Lagrangian algorithm. Due to the impossibility to present all the methods, we shall introduce here two methods within the two above classes. Another class of methods have been formulated recently. These methods are based on the idea to formulate the stress condition in (2.0.1) via a fixed point involving a projection. This is done in [54, 22]. The history of the development of this kind of method is depicted in [55]. In Chapter 4 we shall introduce a more particular projection method that is inherited from image processing [20], that is able to deal with general nonlinearities F.

2.1 Preliminaries

Throughout the next chapters we shall often use some basic results of convex analysis and optimization. This section is dedicated to remind and list several classical results. One can easily find them in [47, 49].

We consider a Hilbert space H, that will be taken later on as either $H = \mathbb{M}^s_{N\times N}(\mathbb{R})$ the space of symmetric square matrices of size N, or an infinite dimensional functional space. The scalar product of two elements $D, \sigma \in H$ will be denoted by $D : \sigma$.

Definition 2.1.1 (Legendre - Fenchel transform)**.** *For $F : H \to \overline{\mathbb{R}}$ proper, we can define $F^* : H \to \overline{\mathbb{R}}$, the conjugate function of F, by*

$$F^*(\sigma) = \sup_{D \in H} (\sigma : D - F(D)), \quad \sigma \in H, \tag{2.1.1}$$

where we denote $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$.

Here the property of F to be proper means that the domain of F, i.e. the set where F has a finite value, is not empty. In other words, F is proper if and only if F is not identically $+\infty$.

Proposition 2.1.2. *Suppose that F is convex, proper and lower semi-continuous (l.s.c.). Then F^* is convex, proper and l.s.c. Moreover $F^{**} = F$.*

We recall the definition of the subdifferential ∂F of the function F, for any $D \in H$,

$$\partial F(D) = \Big\{\sigma \in H \mid F(\overline{D}) \geq F(D) + \sigma : (\overline{D} - D),\ \forall \overline{D} \in H\Big\}. \tag{2.1.2}$$

Lemma 2.1.3. *Let be F a proper, convex, l.s.c. function. Then*
(a) If $\partial F(D)$ is nonempty then $F(D) < \infty$.
(b) For $D, \sigma \in H$,

$$\sigma \in \partial F(D) \Longleftrightarrow D \in \partial F^*(\sigma), \tag{2.1.3}$$

and when this holds one has $F(D) < \infty$ and $F^*(\sigma) < \infty$.
(c) $\sigma \in \partial F(D)$ if and only if the sup in the definition of $F^{**}(D) = F(D)$ is attained at σ.
(d) For $D, \sigma \in H$, one has the Fenchel-Young inequality [49]

$$\sigma : D \leq F(D) + F^*(\sigma). \tag{2.1.4}$$

Moreover, equality holds if and only if $\sigma \in \partial F(D)$.
(e) For $\sigma_1 \in \partial F(D_1)$, $\sigma_2 \in \partial F(D_2)$, we have

$$(\sigma_1 - \sigma_2) : (D_1 - D_2) \geq 0. \tag{2.1.5}$$

(f) $\inf F = -F^*(0)$ *and* $\inf F^* = -F(0)$.

Proof. (a) If $\partial F(D)$ is nonempty, there exists some $\sigma \in \partial F(D)$. Then by definition of the subdifferential,

$$F(\overline{D}) \geq F(D) + \sigma : (\overline{D} - D), \quad \forall \overline{D} \in H. \tag{2.1.6}$$

Since F is proper there is at least one $\overline{D}$ such that $F(\overline{D})$ is finite. Hence the previous inequality yields that $F(D) < \infty$.
(b) Assume that $\sigma \in \partial F(D)$. Then the inequality (2.1.6) holds. According to the definition of F^*, we have

$$F^*(\overline{\sigma}) \geq D : \overline{\sigma} - F(D), \quad \forall \overline{\sigma} \in H. \tag{2.1.7}$$

Using (2.1.6) we deduce that

$$F^*(\overline{\sigma}) \geq D : \overline{\sigma} - F(\overline{D}) + \sigma : (\overline{D} - D), \quad \forall \overline{\sigma} \in H,\ \forall \overline{D} \in H. \tag{2.1.8}$$

Then using the definition of $F^*(\sigma)$ we obtain

$$F^*(\overline{\sigma}) \geq F^*(\sigma) + D : (\overline{\sigma} - \sigma), \quad \forall \overline{\sigma} \in H, \tag{2.1.9}$$

which means that $D \in \partial F^*(\sigma)$. The converse follows from the fact that $F^{**} = F$. Finally the properties $F(D) < \infty$ and $F^*(\sigma) < \infty$ follow from (a).
(c) The definition of $F^{**}(D)$ is

$$F^{**}(D) = \sup_{\overline{\sigma} \in H} (\overline{\sigma} : D - F^*(\overline{\sigma})).$$

To say that the sup is attained at σ means that

$$\overline{\sigma} : D - F^*(\overline{\sigma}) \leq \sigma : D - F^*(\sigma), \quad \forall \overline{\sigma} \in H.$$

Comparing to (2.1.2) this means exactly that $D \in \partial F^*(\sigma)$, or by (b) that $\sigma \in \partial F(D)$.
(d) According to the definition (2.1.1) of F^*, for a given σ one has $F^*(\sigma) \geq \sigma : D - F(D)$ for all D, and the inequality (2.1.4) follows. Equality means that $\overline{D} \mapsto \sigma : \overline{D} - F(\overline{D})$ attains its maximum at D, which by (c) (reverting the role of F and F^*) is equivalent to $D \in \partial F^*(\sigma)$, which gives the claim with the property (b).
(e) When $\sigma_1 \in \partial F(D_1)$ and $\sigma_2 \in \partial F(D_2)$, we have

$$\begin{aligned} F(\overline{D}) &\geq F(D_1) + \sigma_1 : (\overline{D} - D_1), \quad \forall \overline{D} \in H, \\ F(\overline{D}) &\geq F(D_2) + \sigma_2 : (\overline{D} - D_2), \quad \forall \overline{D} \in H. \end{aligned} \tag{2.1.10}$$

Taking $\overline{D} = D_2$ in the first inequality and $\overline{D} = D_1$ in the second one, we get

$$\begin{aligned} F(D_2) &\geq F(D_1) + \sigma_1 : (D_2 - D_1), \\ F(D_1) &\geq F(D_2) + \sigma_2 : (D_1 - D_2). \end{aligned} \tag{2.1.11}$$

Since $\partial F(D_1)$ and $\partial F(D_2)$ are nonempty, we have $F(D_1) < \infty$, $F(D_2) < \infty$. Comparing the two inequalities of (2.1.11) we conclude that $0 \geq \sigma_1 : (D_2 - D_1) + \sigma_2 : (D_1 - D_2)$, which is the claim.
(f) The definition (2.1.1) of F^* gives $F^*(0) = \sup(-F) = -\inf F$, and reversing the role of F and F^* we get the result. □

Proposition 2.1.4 (Moreau envelope and proximal operator)**.** *Let $F : H \to \overline{\mathbb{R}}$ a convex, proper, l.s.c. function. Then for any $\varepsilon > 0$ and $D \in H$ we can define the Moreau envelope of F as*

$$F_\varepsilon(D) := \inf_{\overline{D} \in H} \left\{ F(\overline{D}) + \frac{|\overline{D} - D|^2}{2\varepsilon} \right\}. \tag{2.1.12}$$

Then F_ε is finite everywhere, and
(a) The infimum is attained at the unique point $\widehat{D} = \operatorname{prox}_\varepsilon F(D)$,

$$\operatorname{prox}_\varepsilon F(D) = \operatorname*{argmin}_{\overline{D} \in H} \left\{ F(\overline{D}) + \frac{|\overline{D} - D|^2}{2\varepsilon} \right\}. \tag{2.1.13}$$

Moreover $\widehat{D}$ is characterized by

$$\widehat{D} + \varepsilon \partial F(\widehat{D}) \ni D. \tag{2.1.14}$$

We shall write then

$$\widehat{D} = (\mathrm{Id} + \varepsilon \partial F)^{-1}(D). \tag{2.1.15}$$

(b) The following Moreau identity holds for all $r > 0$ and $\sigma \in H$,

$$(\mathrm{Id} + r\partial F^*)^{-1}(\sigma) + r \left(\mathrm{Id} + \frac{\partial F}{r} \right)^{-1} \left(\frac{\sigma}{r} \right) = \sigma. \tag{2.1.16}$$

(c) The proximity operator $D \mapsto \operatorname{prox}_\varepsilon F(D)$ is 1-Lipschitz continuous.
(d) The Moreau envelope F_ε is locally Lipschitz continuous.
(e) The Moreau envelope F_ε is continuously differentiable with a $1/\varepsilon$ - Lipschitz gradient given for $D \in H$ by

$$F'_\varepsilon(D) = \frac{D - \operatorname{prox}_\varepsilon F(D)}{\varepsilon} \in \partial F(\operatorname{prox}_\varepsilon F(D)). \tag{2.1.17}$$

(f) The Moreau envelope F_ε is convex, and $F_\varepsilon(D) \uparrow F(D)$ pointwise as $\varepsilon \to 0$.

Proof. (a) The proof is classical and can be found in [47, 49].
(b) Denote $\widehat{D} = (\mathrm{Id} + \varepsilon\partial F)^{-1}(D)$. We have

$$\frac{D - \widehat{D}}{\varepsilon} \in \partial F(\widehat{D}) \Longleftrightarrow \widehat{D} \in \partial F^*\left(\frac{D - \widehat{D}}{\varepsilon}\right) \Longleftrightarrow \frac{D}{\varepsilon} \in \frac{1}{\varepsilon}\partial F^*\left(\frac{D - \widehat{D}}{\varepsilon}\right) + \frac{D - \widehat{D}}{\varepsilon}.$$

It follows that

$$\frac{D}{\varepsilon} \in \left(\mathrm{Id} + \frac{\partial F^*}{\varepsilon}\right)\left(\frac{D - \widehat{D}}{\varepsilon}\right) \Longleftrightarrow \frac{D - \widehat{D}}{\varepsilon} = \left(\mathrm{Id} + \frac{\partial F^*}{\varepsilon}\right)^{-1}\left(\frac{D}{\varepsilon}\right).$$

We deduce that

$$D = \varepsilon\left(\mathrm{Id} + \frac{\partial F^*}{\varepsilon}\right)^{-1}\left(\frac{D}{\varepsilon}\right) + \widehat{D} = (\mathrm{Id} + \varepsilon\partial F)^{-1}(D) + \varepsilon\left(\mathrm{Id} + \frac{\partial F^*}{\varepsilon}\right)^{-1}\left(\frac{D}{\varepsilon}\right).$$

Setting $r = 1/\varepsilon$, $D = \varepsilon\sigma$, we obtain (2.1.16).
(c) Denote $\widehat{D}_1 = \operatorname{prox}_\varepsilon F(D_1)$ and $\widehat{D}_2 = \operatorname{prox}_\varepsilon F(D_2)$. Then from (2.1.14) one has

$$\frac{D_1 - \widehat{D}_1}{\varepsilon} \in \partial F(\widehat{D}_1), \qquad \frac{D_2 - \widehat{D}_2}{\varepsilon} \in \partial F(\widehat{D}_2).$$

Using the monotonicity of the subdifferential (2.1.5), one has

$$\begin{aligned}
(D_1 - D_2 - \widehat{D}_1 + \widehat{D}_2) : (\widehat{D}_1 - \widehat{D}_2) &\geq 0,\\
(D_1 - D_2) : (\widehat{D}_1 - \widehat{D}_2) &\geq |\widehat{D}_1 - \widehat{D}_2|^2,\\
|D_1 - D_2||\widehat{D}_1 - \widehat{D}_2| &\geq |\widehat{D}_1 - \widehat{D}_2|^2,\\
|D_1 - D_2| &\geq |\widehat{D}_1 - \widehat{D}_2|.
\end{aligned} \tag{2.1.18}$$

Thus

$$|D_1 - D_2| \geq |\operatorname{prox}_\varepsilon F(\widehat{D}_1) - \operatorname{prox}_\varepsilon F(\widehat{D}_2)|, \tag{2.1.19}$$

which proves the claim.
(d) Inserting $D_2 = 0$, $D_1 = D$ into (2.1.19) gives

$$\begin{aligned}
|D| \geq |\operatorname{prox}_\varepsilon F(D) - \operatorname{prox}_\varepsilon F(0)| &\geq |\operatorname{prox}_\varepsilon F(D)| - |\operatorname{prox}_\varepsilon F(0)|,\\
|D| + |\operatorname{prox}_\varepsilon F(0)| &\geq |\operatorname{prox}_\varepsilon F(D)|.
\end{aligned}$$

Since $\operatorname{prox}_\varepsilon F(0) = \operatorname*{argmin}_{\overline{D}}\left\{F(\overline{D}) + \frac{|\overline{D}|^2}{2\varepsilon}\right\} = C_\varepsilon$, we get $|\operatorname{prox}_\varepsilon F(D)| \leq C_{R,\varepsilon}$ where $|D| \leq R$. Thus $\operatorname{prox}_\varepsilon F(D)$ is bounded when D is bounded. Next, considering $\widehat{D}_1 = \operatorname{prox}_\varepsilon F(D_1)$ and $\widehat{D}_2 =$

$\operatorname{prox}_\varepsilon F(D_2)$, one has

$$\begin{aligned}F_\varepsilon(D_1) &\le F(\widehat{D}_2) + \frac{1}{2\varepsilon}|\widehat{D}_2 - D_1|^2\\ &= F_\varepsilon(D_2) + \frac{1}{2\varepsilon}(|\widehat{D}_2 - D_1|^2 - |\widehat{D}_2 - D_2|^2)\\ &= F_\varepsilon(D_2) + \frac{1}{2\varepsilon}(D_1 + D_2 - 2\widehat{D}_2) : (D_1 - D_2).\end{aligned}$$

Therefore

$$F_\varepsilon(D_1) - F_\varepsilon(D_2) \le \frac{1}{2\varepsilon}(D_1 + D_2 - 2\widehat{D}_2) : (D_1 - D_2) \le \tilde{C}_{R,\varepsilon}|D_1 - D_2|,$$

when D_1, D_2 are in the ball B_R of center 0 and radius R. Similarly, one has

$$F_\varepsilon(D_2) - F_\varepsilon(D_1) \le \tilde{C}_{R,\varepsilon}|D_1 - D_2|,$$

thus for $D_1, D_2 \in B_R$, one has $|F_\varepsilon(D_1) - F_\varepsilon(D_2)| \le \tilde{C}_{R,\varepsilon}|D_1 - D_2|$. This proves that F_ε is locally Lipschitz continuous.
(e) We want to prove that F_ε is differentiable. We consider as above $D_1, D_2 \in H$. Setting

$$M_\varepsilon = \frac{D_1 + D_2 - 2\widehat{D}_1}{2\varepsilon},$$

one has

$$F_\varepsilon(D_2) - F_\varepsilon(D_1) \le \frac{1}{2\varepsilon}(D_1 + D_2 - 2\widehat{D}_1) : (D_2 - D_1) = (D_2 - D_1) : M_\varepsilon.$$

Defining $L(D) = \dfrac{D - \operatorname{prox}_\varepsilon F(D)}{\varepsilon} = \dfrac{D - \widehat{D}}{\varepsilon}$, one has

$$F_\varepsilon(D_2) - F_\varepsilon(D_1) - L(D_1) : (D_2 - D_1) \le (M_\varepsilon - L(D_1)) : (D_2 - D_1) = \frac{|D_2 - D_1|^2}{2\varepsilon}. \tag{2.1.20}$$

Using (2.1.19) we estimate

$$\begin{aligned}&\big(L(D_2) - L(D_1)\big) : (D_1 - D_2)\\ =&\frac{1}{\varepsilon}\big(D_2 - \widehat{D}_2 - D_1 + \widehat{D}_1\big) : (D_1 - D_2)\\ =&-\frac{1}{\varepsilon}|D_1 - D_2|^2 + \frac{1}{\varepsilon}(\widehat{D}_1 - \widehat{D}_2) : (D_1 - D_2)\\ \le&0.\end{aligned} \tag{2.1.21}$$

Reversing the role of D_1 and D_2 in (2.1.20) and using the previous inequality yields

$$F_\varepsilon(D_1) - F_\varepsilon(D_2) - L(D_1) : (D_1 - D_2) \le \frac{|D_2 - D_1|^2}{2\varepsilon}.$$

Hence

$$\Big|F_\varepsilon(D_2) - F_\varepsilon(D_1) - L(D_1) : (D_2 - D_1)\Big| \le \frac{|D_2 - D_1|^2}{2\varepsilon}. \tag{2.1.22}$$

We deduce that F_ε is differentiable and $F'_\varepsilon(D) = L(D) = \dfrac{D - \operatorname{prox}_\varepsilon F(D)}{\varepsilon}$. Using (2.1.18) we write

$$(D_1 - D_2 - \widehat{D}_1 + \widehat{D}_2) : (\widehat{D}_1 - \widehat{D}_2 - D_1 + D_2) \ge (D_1 - D_2 - \widehat{D}_1 + \widehat{D}_2) : (D_2 - D_1),$$

thus

$$|D_1 - D_2 - \widehat{D}_1 + \widehat{D}_2|^2 \leq (D_1 - D_2 - \widehat{D}_1 + \widehat{D}_2) : (D_2 - D_2),$$

which implies that

$$|D_1 - D_2 - \widehat{D}_1 + \widehat{D}_2| \leq |D_1 - D_2|. \tag{2.1.23}$$

This proves that $\mathrm{Id} - \mathrm{prox}_\varepsilon F$ is 1-Lipschitz continuous. Therefore F'_ε is $1/\varepsilon$-Lipschitz continuous.
(f) The inequality (2.1.21) implies that $(F'_\varepsilon(D_2) - F'_\varepsilon(D_1)) : (D_2 - D_1) \geq 0$, which gives that F_ε is convex. Then, the definition (2.1.12) gives directly that F_ε increases as ε decreases. Taking $\overline{D} = D$ we have also obviously that $F_\varepsilon(D) \leq F(D)$. Then, for a fixed D, since F is l.s.c., for any $\lambda < F(D)$ there is a ball $B(D, r)$ around D in which $F \geq \lambda$. Then for $\overline{D} \in B(D, r)$ one has $F(\overline{D}) + |\overline{D} - D|^2/2\varepsilon \geq \lambda$. But F is lower bounded by a linear function, $F(\overline{D}) \geq A + \sigma_0 : \overline{D}$ for some real number A and $\sigma_0 \in H$. Thus for $\overline{D} \notin B(D, r)$,

$$\begin{aligned}
&F(\overline{D}) + |\overline{D} - D|^2/2\varepsilon \\
\geq &|\overline{D} - D|^2/2\varepsilon + A + \sigma_0 : D - |\sigma_0||\overline{D} - D| \\
\geq &|\overline{D} - D|^2/4\varepsilon + A + \sigma_0 : D - \varepsilon|\sigma_0|^2 \\
\geq &r^2/4\varepsilon + A + \sigma_0 : D - \varepsilon|\sigma_0|^2.
\end{aligned}$$

Thus for ε small enough this will be larger than λ, and it follows that $F_\varepsilon(D) \geq \lambda$, which finishes the proof that $F_\varepsilon(D)$ tends to $F(D)$ as ε tends to 0. □

Remark: Any convex, proper and lower semi-continuous F is lower bounded by an affine function. It follows from Proposition 2.1.4(a): by taking arbitrary $D \in H$ and $\varepsilon > 0$ we get some $\widehat{D} \in H$ satisfying (2.1.14), which implies that $\partial F(\widehat{D}) \neq \emptyset$. Thus there is some $\sigma \in \partial F(\widehat{D})$, and the definition (2.1.2) of $\partial F(\widehat{D})$ ensures that F is lower bounded by the affine function $\overline{D} \mapsto F(\widehat{D}) + \sigma : (\overline{D} - \widehat{D})$.
Remark: Since F is convex, proper and lower semi-continuous, it follows that F is continuous on the interior of its domain [49].

2.2 The regularization approach

One of the intriguing obstacles in the modelling of viscoplastic media is the presence of unknown interfaces separating the yielded and the un-yielded regions, that are difficult to track. In (2.0.1) they correspond to locations where the stress $\sigma \in \partial F(Du)$ switches from a place where F is differentiable to one that is not. To avoid the associated numerical difficulties, the regularization method replaces the nonlinearity F by a smooth one, with a small approximation parameter ε. Then the material behaves as a fluid in the entire computational domain, and we have a differentiable, convex, proper, l.s.c .function F_ε. Note that here F_ε can be any function that approximates F. It can be the Moreau envelope of F defined by (2.1.12), or any other approximation. The approximate system is then

$$\begin{cases} \alpha u - \operatorname{div} F'_\varepsilon(Du) = f & \text{in} \quad \Omega, \\ u = 0 & \text{on} \quad \Gamma. \end{cases}$$

This system can be solved by an iterative method with linearization, which is possible since F_ε is smooth. The classical theory of monotone operators is developed in [16, 58, 25]. The solution to

the former problem can be formulated via a minimization problem as

$$\inf_{u\in H}\int_\Omega \left(\alpha\frac{|u|^2}{2}+F(Du)-f\cdot u\right)\mathrm{dx},$$

where H is an appropriate Hilbert space of functions $u(x)$ such that $u\mapsto\int F(Du)$ is convex, proper and l.s.c. There is here a difficulty to define such space for a general nonlinearity F, in particular concerning the lower semi-continuity. This will be studied in Chapter 3 for quite general nonlinearities F. In the regularization approach we have the approximate differentiable nonlinear optimization problem

$$\inf_{u_\varepsilon\in H}\int_\Omega \left(\alpha\frac{|u_\varepsilon|^2}{2}+F_\varepsilon(Du_\varepsilon)-f\cdot u_\varepsilon\right)\mathrm{dx}\,.$$

In fact the regularization approach has until now only been considerd for particular nonlinearities F, that is for the Herschel-Bulkley model or the Bingham model, with or without viscosity. Let us here consider the Bingham fluid with, which corresponds to $F(D)=\sigma_0|D|+\eta|D|^2/2$, for some yield stress $\sigma_0>0$ and viscosity $\eta\geq 0$. This particular problem can be considered as a model for quasi-Newtonian viscous fluids [53, Chapter 2]. Then we have to find $u\in H$ such that

$$\alpha u-\eta\,\mathrm{div}\,Du-\sigma_0\,\mathrm{div}\left(\frac{Du}{|Du|}\right)=f\quad\text{in}\quad\Omega, \tag{2.2.1}$$

or

$$\inf_{u\in H}\int_\Omega \left(\alpha\frac{|u|^2}{2}+\eta\frac{|Du|^2}{2}+\sigma_0|Du|-f\cdot u\right)\mathrm{dx}\,.$$

In the approximate problem we look for $u_\varepsilon\in H$ realizing

$$\inf_{u_\varepsilon\in H}\int_\Omega \left(\alpha\frac{|u_\varepsilon|^2}{2}+\eta\frac{|Du_\varepsilon|^2}{2}+\sigma_0(\sqrt{|Du_\varepsilon|^2+\varepsilon^2}-\varepsilon)-f\cdot u_\varepsilon\right)\mathrm{dx},$$

where $\varepsilon>0$. It corresponds to solve the well-posed nonlinear parabolic problem

$$\alpha u_\varepsilon-\eta\,\mathrm{div}\,Du_\varepsilon-\sigma_0\,\mathrm{div}\,\frac{Du_\varepsilon}{\sqrt{|Du_\varepsilon|^2+\varepsilon^2}}=f\quad\text{in}\quad\Omega. \tag{2.2.2}$$

It can be solved by an iterative algorithm,

$$\alpha u_\varepsilon^{k+1}-\eta\,\mathrm{div}\,Du_\varepsilon^{k+1}-\sigma_0\,\mathrm{div}\,\frac{Du_\varepsilon^{k+1}}{\sqrt{|Du_\varepsilon^k|^2+\varepsilon^2}}=f\quad\text{in}\quad\Omega, \tag{2.2.3}$$

each iteration giving rise to the resolution of a linear problem. An overall space discretization has to be performed, usually by finite elements. The regularization method has certain advantages, namely it can use classical numerical schemes and it can be implemented easily in many existing codes. One of the key points in this method is finding the optimal value for ε. Indeed the smaller is ε the smaller is the error done in the law (one can prove that $\|u-u_\varepsilon\|_{L^2}\simeq\sqrt{\varepsilon}$), but larger is the number of iterations needed. Thus for a given tolerance one has to choose the corresponding ε, as well as the corresponding space size Δx, as usual. Therefore some optimal relation between ε and Δx has to be found. In [14] the authors propose an optimal relation of the form $\varepsilon\sim(\Delta x)^2$ for P^1 finite elements (or $\varepsilon\sim(\Delta x)^4$ in some cases when the solution less regular). A good formula for the choice of ε is

$$\varepsilon\sim 10^{-2}\Delta x^2|Du|^2\frac{\alpha}{\sigma_0}, \tag{2.2.4}$$

in the case when the viscosity is small. Most of the time the regularization runs quite fast.

There are many studies using this classical approach, as [7, 21, 46]. Some studies [42, 38] depicted certain drawbacks of this method: the yield/plug zones (where $Du = 0$) can be not well resolved, or the asymptotic behaviour as $t \to \infty$ can be wrong. This is the motivation for developing alternative approaches.

2.3 Augmented Lagrangian method

The second approach, especially designed to overcome the difficulty of non-differentiability of the constitutive law, strongly uses the theory of variational inequalities that one can find in Duvaut and Lions [25]. The viscoplastic problem can be formulated as a saddle point problem with Lagrange multipliers. Then the well-known Augmented Lagrangian method for solving constrained optimization problems, first introduced by Fortin and Glowinski [28], can be applied. It is also is called ADMM algorithm (Alternate directions method of multipliers), see [20].

Recall first that the viscoplastic problem (2.0.1) can be formulated via the convex-concave Lagrangian

$$\mathcal{L}(\sigma, u) = \int_\Omega \left(\alpha \frac{|u|^2}{2} + \sigma : Du - F^*(\sigma) - f \cdot u \right) \mathrm{dx}, \tag{2.3.1}$$

as the two equivalent problems

$$\inf_u \sup_\sigma \mathcal{L}(\sigma, u) = \sup_\sigma \inf_u \mathcal{L}(\sigma, u). \tag{2.3.2}$$

The principle of the method is to introduce an additional variable γ that is an approximation of Du, and to set the Lagrangian multiplier corresponding to the constrain $\gamma - Du = 0$. Recalling that $F^*(\sigma) = \sup_\gamma(\sigma : \gamma - F(\gamma))$, the problem (2.3.2) can be rewritten as

$$\sup_\sigma \inf_{u,\gamma} \overline{\mathcal{L}}(\sigma, u, \gamma), \tag{2.3.3}$$

with

$$\overline{\mathcal{L}}(\sigma, u, \gamma) = \int_\Omega \left(\alpha \frac{|u|^2}{2} + \sigma : (Du - \gamma) + F(\gamma) - f \cdot u \right) \mathrm{dx}\,. \tag{2.3.4}$$

The augmented formulation is to rewrite the problem as

$$\sup_\sigma \inf_{u,\gamma} \overline{\mathcal{L}}^{aug}(\sigma, u, \gamma), \tag{2.3.5}$$

with the augmented Lagrangian

$$\overline{\mathcal{L}}^{aug}(\sigma, u, \gamma) = \overline{\mathcal{L}}(\sigma, u, \gamma) + \frac{r}{2} \|Du - \gamma\|_{L^2}^2, \tag{2.3.6}$$

where $r > 0$ is a free parameter. The augmented term does not change the minimality condition, nor the value of the minimum, but enforces the stability of the numerical methods. The algorithm proposed by Fortin and Glowinski [28] contains the following steps:

- Initialize with given σ_0, γ_0.
- For $k \geq 0$, assuming that σ_k, γ_k are known, calculate u_k by $u_k := \operatorname{argmin}_u \overline{\mathcal{L}}^{aug}(\sigma_k, u, \gamma_k)$, i.e. u_k solves

$$\alpha u_k - \operatorname{div} \sigma_k - \operatorname{div} r(Du_k - \gamma_k) = f. \tag{2.3.7}$$

- Calculate γ_{k+1} by $\gamma_{k+1} := \underset{\gamma}{\operatorname{argmin}}\, \overline{\mathcal{L}}^{aug}(\sigma_k, u_k, \gamma)$, i.e.

$$\partial F(\gamma_{k+1}) \ni \sigma_k - r(\gamma_{k+1} - Du_k). \tag{2.3.8}$$

- Calculate σ_{k+1} by going into the direction of the gradient of $\overline{\mathcal{L}}^{aug}$ with respect to σ, i.e.

$$\sigma_{k+1} = \sigma_k + r(Du_k - \gamma_{k+1}). \tag{2.3.9}$$

The first step consists in solving an elliptic problem, which can be done by the finite elements method (FEM). The two latter steps are nonlinear but local, they can be calculated pointwise efficiently. With the definition (2.4.1) of $\mathbb{P}_r$ and the Moreau identity (2.1.16), the update of γ and σ can be written equivalently as

$$\gamma_{k+1} = (\mathrm{Id} + \partial F/r)^{-1}\left(\frac{\sigma_k + rDu_k}{r}\right) = \frac{\sigma_k + rDu_k - \sigma_{k+1}}{r}, \qquad \sigma_{k+1} = \mathbb{P}_r(\sigma_k + rDu_k). \tag{2.3.10}$$

The Augmented Lagrangian method was introduced in the 70s but is still used, one can find it in the works of Saramito et al. [15]. However, the rate of convergence is still limited in comparison with the regularization method when accurate determination of solid zones is not required. Another delicate issue in this method is the determination of an appropriate value of r in order to have a good convergence rate. An improvement with this respect is the Bermudez-Moreno scheme. Besides, the preconditioned version of the ADMM method is equivalent to the primal-dual algorithm proposed in [20], which is studied in Chapter 4.

2.4 Projection approach

We would like here to introduce a class of methods that have a formulation which is close to the Augmented Lagrangian method, retaining the essential projection step (2.3.10). In particular, the algorithms used in imaging such as [20] fall into this class. The most challenging task when solving the viscoplastic model (2.0.1) is to solve the constraint $\sigma \in \partial F(Du)$, since a single value of Du leads to several possible values of σ. In the projection approach, the main idea is to formulate the constraint via a fixed point on σ involving a projection. This kind of formulation is used in [22, 19] in the context of incompressible Navier-Stokes equations with a viscous Bingham viscoplastic law. In the incompressible setting, this approach is well suited in conjunction with the projection on free divergence vector fields, that is described in [31] and that is useful for the simulation of turbulent flows. The approach of [22] relies strongly on the presence of viscosity and can be interpreted as a time regularization.

In this subsection, a projection formulation is introduced for a quite general nonlinearity F, and viscosity is not necessary.

As in Section 2.1 we consider a convex, proper and l.s.c. function F defined on a Hilbert space H, that will be taken later on as either the space $\mathbb{M}^s_{N\times N}(\mathbb{R})$ of symmetric square matrices of size N or eventually a space of functions. Then the proximal operator of F^*, as defined in Proposition 2.1.4, is well defined. We consider thus for $r > 0$ the operator $\mathbb{P}_r$ defined as

$$\mathbb{P}_r(\sigma) = (\mathrm{Id} + r\partial F^*)^{-1}(\sigma), \quad \sigma \in H. \tag{2.4.1}$$

Then according to Proposition 2.1.4(c), $\mathbb{P}_r$ is $1-$Lipschitz continuous.

Lemma 2.4.1. *For $r > 0$ and $\mathbb{P}_r$ defined as (2.4.1), one has:*
(a) For $\sigma \in H$,

$$\mathbb{P}_r(\sigma) = F_\varepsilon{}'\left(\frac{\sigma}{r}\right), \quad \text{with } \varepsilon = \frac{1}{r}, \tag{2.4.2}$$

where F_ε is the Moreau envelope of F.
(b) $\mathbb{P}_r$ is $1-$Lipschitz continuous, and monotone i.e. for any $\sigma_1, \sigma_2 \in H$

$$\big(\mathbb{P}_r(\sigma_2) - \mathbb{P}_r(\sigma_1)\big) : (\sigma_2 - \sigma_1) \geq 0. \tag{2.4.3}$$

(c) For $D, \sigma \in H$, one has the equivalence

$$\sigma \in \partial F(D) \Leftrightarrow \mathbb{P}_r(\sigma + rD) = \sigma. \tag{2.4.4}$$

(d) $\sigma \in \partial F(0) \Leftrightarrow \mathbb{P}_r(\sigma) = \sigma$.

Proof. It follows from Proposition 2.1.4.
(a) The formula (2.1.17) applied to $D = \sigma/r$ and the Moreau identity (2.1.16) yield (2.4.2).
(b) Proposition 2.1.4(f) gives the monotonicity.
(c) Since $\mathbb{P}_r(\sigma + rD) = (\mathrm{Id} + r\partial F^*)^{-1}(\sigma + rD)$, we have

$$\mathbb{P}_r(\sigma + rD) = \sigma \Longleftrightarrow \sigma + rD \in (\mathrm{Id} + r\partial F^*)(\sigma) \Longleftrightarrow D \in \partial F^*(\sigma) \Longleftrightarrow \sigma \in \partial F(D).$$

(d) It follows from (c) by taking $D = 0$. □

The previous lemma enables to formulate the constraint $\sigma \in \partial F(D)$ of (2.0.1) as σ being a fixed point of the map $\sigma \mapsto \mathbb{P}_r(\sigma + rD)$, which is a $1-$Lipschitz mapping. Then one can think of an iteration procedure in order to get a solution σ. However the mapping is not a contraction (which is related to the fact that there could be several solutions), and thus the convergence is slow, if it holds. This is studied in Chapter 4 when the iteration is coupled with the momentum equation. Indeed according to Lemma 2.4.1, the viscoplastic model (2.0.1) can be rewritten as

$$\begin{cases} \alpha u - \operatorname{div} \sigma = f, \\ \sigma = \mathbb{P}_r(\sigma + rDu), \end{cases} \tag{2.4.5}$$

where r is any positive constant. Note that the set $\partial F(0)$ plays a particular role here since $\sigma \in \partial F(0)$ corresponds to the stresses σ that are admissible in solid zones i.e. where $Du = 0$.

An issue is how to compute $\mathbb{P}_r$ if F is known (but F^* is not known explicitly). One can think of using the Moreau identity (2.1.16). When F has some regularity outside the origin we can use the following formulas.

Proposition 2.4.2. *Assume that F is a convex, proper, l.s.c. function on H, and that*

$$F \text{ is finite everywhere, and } F \text{ is differentiable outside the origin.} \tag{2.4.6}$$

Then for $\sigma \notin \partial F(0)$, there exists a unique $D \neq 0$ such that $F'(D) + rD = \sigma$. For $r > 0$ we then have

$$\mathbb{P}_r(\sigma) = \begin{cases} \sigma & \text{if } \sigma \in \partial F(0), \\ F'((F' + r\,\mathrm{Id})^{-1}(\sigma)) & \text{if } \sigma \notin \partial F(0). \end{cases} \tag{2.4.7}$$

Proof. According to Moreau's identity (2.1.16), we have

$$\mathbb{P}_r(\sigma) = (\mathrm{Id} + r\partial F^*)^{-1}(\sigma) = \sigma - r\left(\mathrm{Id} + \frac{\partial F}{r}\right)^{-1}\left(\frac{\sigma}{r}\right). \tag{2.4.8}$$

For the case $\sigma \in \partial F(0)$, we have $\left(\mathrm{Id} + \frac{\partial F}{r}\right)(0) \ni \frac{\sigma}{r}$, thus $\left(\mathrm{Id} + \frac{\partial F}{r}\right)^{-1}\left(\frac{\sigma}{r}\right) = 0$. Inserting it into (2.4.8) we obtain $\mathbb{P}_r(\sigma) = \sigma$.
For $\sigma \notin \partial F(0)$ there exists a unique D such that $\partial F(D) + rD \ni \sigma$. Then since $\sigma \notin \partial F(0)$, D cannot be zero, and thus $F'(D) + rD = \sigma$. It implies

$$\left(\mathrm{Id} + \frac{\partial F}{r}\right)^{-1}\left(\frac{\sigma}{r}\right) = D.$$

Hence we get

$$F'((F' + r\,\mathrm{Id})^{-1}(\sigma)) + r\left(\mathrm{Id} + \frac{\partial F}{r}\right)^{-1}\left(\frac{\sigma}{r}\right) = F'(D) + rD = \sigma.$$

Therefore inserting it into (2.4.8), for $\sigma \notin \partial F(0)$ we conclude that $\mathbb{P}_r(\sigma) = F'((F' + r\,\mathrm{Id})^{-1}(\sigma))$. □

2.5 Pure plastic models

We provide here particular properties for pure plastic models, which means that F is homogeneous of degree 1. The mechanical interpretation is that in this case the law does not include any viscous effect. These models are also called "rate independent" because when $\sigma \in \partial F(D)$, σ depends only on the direction of D, but not on its magnitude.

Proposition 2.5.1. *Suppose that F is a convex, l.s.c, proper function on H. Then F is homogeneous of degree 1 i.e.*

$$F(\lambda D) = \lambda F(D), \quad \textit{for all } \lambda > 0 \textit{ and } D \in H, \tag{2.5.1}$$

if and only if there exists $\mathcal{A} \subset H$ convex, closed and nonempty such that

$$F^*(\sigma) = \begin{cases} 0 & \textit{if } \sigma \in \mathcal{A}, \\ \infty & \textit{if } \sigma \notin \mathcal{A}. \end{cases} \tag{2.5.2}$$

Moreover, we have then

$$F(D) = \sup_{\sigma \in \mathcal{A}} \sigma : D. \tag{2.5.3}$$

Proof. Let us start from recalling the Fenchel conjugate function $F^*(\sigma) = \sup_D(\sigma : D - F(D))$. If F is homogeneous, then for all $\sigma \in H$ and $\lambda > 0$, we have

$$\begin{aligned} F^*(\lambda\sigma) &= \sup_D(\lambda\sigma : D - F(D)) = \sup_D\left(\lambda\sigma : \frac{D}{\lambda} - F\left(\frac{D}{\lambda}\right)\right) = \sup_D\left(\sigma : D - \frac{F(D)}{\lambda}\right) \\ &= \frac{1}{\lambda}\sup_D(\lambda\sigma : D - F(D)) = \frac{1}{\lambda}F^*(\lambda\sigma). \end{aligned}$$

Taking $\lambda = 2$ we deduce that $F^*(2\sigma) = 0$ or ∞, for all σ. It means that $F^*(\sigma) = 0$ or ∞ for all σ. Hence F^* takes the form

$$F^*(\sigma) = \begin{cases} 0 & \text{if } \sigma \in \mathcal{A}, \\ \infty & \text{if } \sigma \notin \mathcal{A}, \end{cases}$$

for some set $\mathcal{A} \subset H$. We know that F^* is convex, l.s.c, and proper. Since F^* is proper it implies that $\mathcal{A} \neq \emptyset$. Since F^* is convex we get that $\mathcal{A}$ is necessarily convex. Lastly, to prove that $\mathcal{A}$ is closed, take an arbitrary sequence $\sigma_n \in \mathcal{A}$ converging to $\sigma \in H$. Because F^* is lower semi-continuous, $F^*(\sigma) \leq \underline{\lim} F^*(\sigma_n) = 0$. Thus $F^*(\sigma) = 0$ and $\sigma \in \mathcal{A}$. It proves that $\mathcal{A}$ is closed.
Conversely, if we define F^* by (2.5.2) for some $\mathcal{A}$ convex, closed and non-empty, as previously we have that F^* is convex, lower semi-continuous and proper. Then $F(D) = F^{**}(D) = \sup_\sigma(D : \sigma - F^*(\sigma)) = \sup_{\sigma \in \mathcal{A}}(\sigma : D)$. For all $\lambda > 0, D \in H$, we have $F(\lambda D) = \sup_{\sigma \in \mathcal{A}}(\lambda\sigma : D) = \lambda \sup_{\sigma \in \mathcal{A}}(\sigma : D) = \lambda F(D)$. Hence F is homogeneous of degree 1. Moreover the formula (2.5.3) is proved. □

Lemma 2.5.2. *Suppose that F is convex, l.s.c, proper, and homogeneous of degree 1. Then we have the following properties:*
(a) $F(0) = 0$.
(b) The set $\mathcal{A}$ in (2.5.2) is given by $\mathcal{A} = \partial F(0)$.
(c) $\mathcal{A}$ is bounded if and only if $\operatorname{dom} F = H$.

Proof. (a) Taking $D = 0$, $\lambda = 2$ in (2.5.1), we deduce $F(0) = 2F(0)$. Thus $F(0) = 0$ or $F(0) = +\infty$. If $F(0) < \infty$ then we are done. Otherwise, since F is proper, there exists $D \neq 0$ such that $F(D) < \infty$. Since F is l.s.c, we have

$$F(0) \leq \underline{\lim}_{\lambda \to 0^+} F(\lambda D) = \underline{\lim}_{\lambda \to 0^+} \lambda F(D) = 0.$$

This implies that $F(0) = 0$.
(b) We have the equivalences

$$\begin{aligned} \sigma \in \mathcal{A} &\Longleftrightarrow F^*(\sigma) \leq 0 \Longleftrightarrow \sup_D(\sigma : D - F(D)) \leq 0 \Longleftrightarrow F(D) \geq \sigma : D \quad \forall D \\ &\Longleftrightarrow F(D) - F(0) \geq \sigma : (D - 0) \quad \forall D \Longleftrightarrow \sigma \in \partial F(0). \end{aligned}$$

Hence $\mathcal{A} = \partial F(0)$.
(c) Because of (2.5.3), the property that $F(D) < \infty$ for all $D \in H$ means that $\mathcal{A}$ is weakly bounded, which is equivalent to $\mathcal{A}$ bounded by the Banach-Steinhaus theorem. □

Proposition 2.5.3. *Suppose that F^* is defined as (2.5.2) for some $\mathcal{A}$ convex, closed and nonempty. Then we have:*
(a) If $\sigma \in \operatorname{int}(\mathcal{A})$, then $\partial F^*(\sigma) = \{0\}$.
(b) If $\sigma \notin \mathcal{A}$, then $\partial F^*(\sigma) = \emptyset$.
(c) If $\sigma \in \partial\mathcal{A}$, then

$$D \in \partial F^*(\sigma) \quad \Longleftrightarrow \quad D : (\tau - \sigma) \leq 0 \quad \forall \tau \in \mathcal{A}. \tag{2.5.4}$$

*When D satisfies (2.5.4), we say that D is an **outer direction** of $\mathcal{A}$ at $\sigma \in \partial\mathcal{A}$.*
(d) In the case of a finite dimensional space H, if $\mathcal{A}$ has a C^1 boundary and $\operatorname{int}\mathcal{A} \neq \emptyset$, then for $\sigma \in \partial\mathcal{A}$,

$$\partial F^*(\sigma) = \mathbb{R}_+ \, \boldsymbol{n}(\sigma),$$

where $\mathrm{n}(\sigma)$ is the exterior normal of $\partial\mathcal{A}$ at the point σ.

Proof. (a) Assume that $\sigma \in \text{int}(\mathcal{A})$. Since $F^* \geq 0$ and $F^*(\sigma) = 0$, it follows that $0 \in \partial F^*(\sigma)$. Conversely suppose that $D \in \partial F^*(\sigma)$. Then there exists $\varepsilon > 0$ such that $\sigma + \varepsilon D \in \mathcal{A}$. It follows that $F^*(\sigma + \varepsilon D) = 0$. Since $D \in \partial F^*(\sigma)$,

$$F^*(\sigma + \varepsilon D) \geq F^*(\sigma) + D : (\sigma + \varepsilon D - \sigma),$$

thus $0 \geq \varepsilon |D|^2$ and $D = 0$. Therefore $\partial F^*(\sigma) \subset \{0\}$.
(b) If $\sigma \notin \mathcal{A}$, we have $F^*(\sigma) = \infty$, and by Lemma 2.1.3(a) it follows that $\partial F^*(\sigma) = \emptyset$.
(c) If $\sigma \in \partial\mathcal{A}$, a vector $D \in H$ verifies $D \in \partial F^*(\sigma)$ if and only if

$$\begin{aligned} F^*(\tau) &\geq F^*(\sigma) + D : (\tau - \sigma) \quad \forall \tau, \\ 0 &\geq D : (\tau - \sigma) \quad \forall \tau \in \mathcal{A}. \end{aligned}$$

This proves the claim.
(d) The case of C^1 boundary is obvious and not detailed here. □

Remark: Some references call a function of the form (2.5.2) the indicator of the closed convex set $\mathcal{A}$, denoted by $\text{Id}_{\mathcal{A}}$. And the subdifferential $\partial F^*(\sigma)$ of an indicator of a closed convex set $\mathcal{A}$, for $\sigma \in \partial\mathcal{A}$, is called a "normal cone".

Proposition 2.5.4. *Suppose that F is convex, l.s.c, proper and homogeneous of degree 1.*
(a) If $D = 0$, then $\partial F(D) = \mathcal{A}$.
(b) If $D \neq 0$, then $\partial F(D) = \{\sigma \in \partial\mathcal{A} \mid D$ is an outer direction of $\mathcal{A}$ at $\sigma\}$.

Proof. It follows from Proposition 2.5.3 by using the fact that $D \in \partial F^*(\sigma) \Longleftrightarrow \sigma \in \partial F(D)$. □

Remark: In the case of C^1 boundary as stated in Proposition 2.5.3(d), when we have the relation $\sigma \in \partial F(D)$ and D varies ($D \neq 0$), σ remains in $\partial\mathcal{A}$. Then a small variation δD of D induces a small variation $\delta\sigma$ of σ, that verifies $\delta\sigma : D = 0$. In such situation, mechanics call the law F "associated". Slightly different presentations of pure plastic models are possible, see [17].

Theorem 2.5.5. *Suppose that F is convex, l.s.c, proper and homogeneous of degree 1. Let $\mathcal{A} = \partial F(0)$. If we define $\mathbb{P}_r(\sigma) = (\text{Id} + r\partial F^*)^{-1}(\sigma)$ for $r > 0$, then $\mathbb{P}_r(\sigma)$ is independent of r and is the orthogonal projection of σ on $\mathcal{A}$. To find the projection $\mathbb{P}_r(\sigma)$ when $\sigma \notin \mathcal{A}$, we have to find $\sigma_b \in \partial\mathcal{A}$ and n_b an outer direction of $\mathcal{A}$ at σ_b, such that*

$$\sigma = \sigma_b + n_b. \tag{2.5.5}$$

Then $\mathbb{P}_r(\sigma) = \sigma_b$.

Proof. We recall that for a nonempty closed convex set $\mathcal{A}$ we can define uniquely an orthogonal projection $\mathbb{P}_{\mathcal{A}}$ on $\mathcal{A}$ such that for any $\sigma \in H$, $\mathbb{P}_{\mathcal{A}}(\sigma)$ is characterized by

$$(\sigma' - \mathbb{P}_{\mathcal{A}}(\sigma)) : (\sigma - \mathbb{P}_{\mathcal{A}}(\sigma)) \leq 0 \quad \forall \sigma' \in \mathcal{A}. \tag{2.5.6}$$

By definition of the projection $\mathbb{P}_r$, there exists some $D \in H$ such that

$$\mathbb{P}_r(\sigma) + rD = \sigma \text{ with } \mathbb{P}_r(\sigma) \in \partial F(D). \tag{2.5.7}$$

Since $D \in \partial F^*(\mathbb{P}_r(\sigma))$, we have that $\partial F^*(\mathbb{P}_r(\sigma)) \neq \emptyset$ and that by Proposition 2.5.3 we obtain $P_r(\sigma) \in \mathcal{A}$. Thus either $\mathbb{P}_r(\sigma) \in \text{int}(\mathcal{A})$ or $\mathbb{P}_r(\sigma) \in \partial\mathcal{A}$.

If $\mathbb{P}_r(\sigma) \in \mathrm{int}(\mathcal{A})$ then $D = 0$ and $\mathbb{P}_r(\sigma) = \sigma = \mathbb{P}_{\mathcal{A}}(\sigma)$.
If $\mathbb{P}_r(\sigma) \in \partial\mathcal{A}$, since $D \in \partial F^*(\mathbb{P}_r(\sigma))$ we have according to Proposition 2.5.3(c)

$$0 \geq D : (\sigma' - \mathbb{P}_r(\sigma)) \quad \forall \sigma' \in \mathcal{A},$$
$$0 \geq \frac{\sigma - \mathbb{P}_r(\sigma)}{r} : (\sigma' - \mathbb{P}_r(\sigma)) \quad \forall \sigma' \in \mathcal{A}.$$

This means that $\mathbb{P}_r(\sigma)$ is the orthogonal projection of σ on the convex set $\mathcal{A}$. This proves that $\mathbb{P}_r = \mathbb{P}_{\mathcal{A}}$.
For the formula (2.5.5), consider $\sigma \notin \mathcal{A}$. From the definition one has $\mathbb{P}_r(\sigma) = (\mathrm{Id} + r\partial F^*)^{-1}(\sigma)$, which means that

$$\mathbb{P}_r(\sigma) + r\partial F^*(\mathbb{P}_r(\sigma)) \ni \sigma. \tag{2.5.8}$$

It follows that $\partial F^*(\mathbb{P}_r(\sigma)) \neq \emptyset$. By Proposition 2.5.3 and since $\sigma \notin \mathcal{A}$ we get that $\mathbb{P}_r(\sigma) \in \partial\mathcal{A}$. Then (2.5.8) means that $\sigma - \mathbb{P}_r(\sigma)$ is an outer direction of $\mathcal{A}$ at $\mathbb{P}_r(\sigma)$, which proves (2.5.5). □

2.6 Examples

In this subsection we apply the results of the previous section to various pure plastic models, corresponding to different choices of $F(D)$ defined on the space $H = \mathbb{M}^s_{N\times N}(\mathbb{R})$ of symmetric square matrices of size N.

Bingham model

In this model we take $F(D) = |D|$. This is the most simple choice. Using that $\mathcal{A} = \partial F(0)$, we obtain that $\sigma \in \mathcal{A}$ is characterized by

$$\sigma : D \leq |D| \quad \forall D.$$

It follows that $\mathcal{A} = \{\sigma : |\sigma| \leq 1\}$. We can represent $\mathcal{A}$ geometrically as a half circle in the variables p, $|\sigma'|$ with $\sigma = -p\,\mathrm{Id} + \sigma'$, $\mathrm{Tr}\,\sigma' = 0$, as shown on Figure 2.1a. In [22], L. Chupin and T. Dubois used the associated projection $\mathbb{P}_{\mathcal{A}}$ on $\mathcal{A}$. It is given simply by

$$\mathbb{P}_{\mathcal{A}}(\sigma) = \begin{cases} \sigma & \text{if } |\sigma| \leq 1, \\ \dfrac{\sigma}{|\sigma|} & \text{if } |\sigma| > 1. \end{cases} \tag{2.6.1}$$

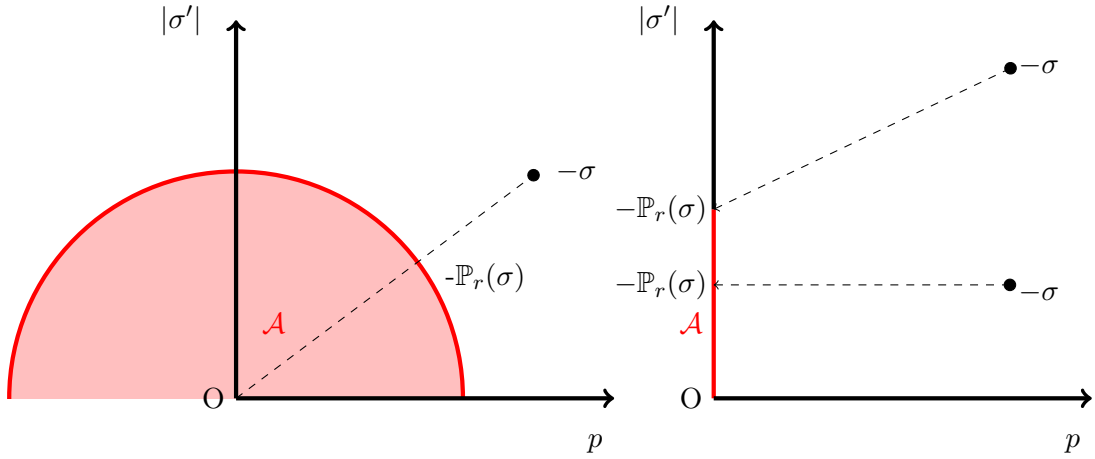
Degenerate Bingham model

In this model we take $F(D) = |D'|$, with D' the deviator of D, i.e. $D' = D - \mathrm{Tr}(D)\,\mathrm{Id}/N$. Then since $\mathcal{A} = \partial F(0)$, the stresses $\sigma \in \mathcal{A}$ are characterized by

$$\sigma : D \leq |D'| \quad \forall D, \tag{2.6.2}$$
$$\left(\mathrm{Tr}(\sigma)\frac{\mathrm{Id}}{N} + \sigma'\right) : \left(\mathrm{Tr}(D)\frac{\mathrm{Id}}{N} + D'\right) \leq |D'| \quad \forall D, \tag{2.6.3}$$
$$\frac{\mathrm{Tr}(\sigma)\,\mathrm{Tr}(D)}{N} + \sigma' : D' \leq |D'| \quad \forall D. \tag{2.6.4}$$

It implies that $\mathcal{A} = \{\sigma = -p\,\mathrm{Id} + \sigma' \mid p = 0, |\sigma'| \leq 1\}$. Next we compute the projection $\mathbb{P}_r(\sigma)$ from (2.4.1), (2.4.8),

$$\mathbb{P}_r(\sigma) = \sigma - r\left(\mathrm{Id} + \frac{\partial F}{r}\right)^{-1}\left(\frac{\sigma}{r}\right). \tag{2.6.5}$$



(a) Bingham model

(b) Degenerate Bingham model

Figure 2.1: Set $\mathcal{A}$ and orthogonal projection

Let us denote $D = \left(\mathrm{Id} + \frac{\partial F}{r}\right)^{-1}\left(\frac{\sigma}{r}\right)$. Then $\sigma \in rD + \partial F(D)$, which means that

$$\begin{cases} \sigma = rD + \dfrac{D'}{|D'|} & \text{if} \quad |D'| \neq 0, \\ \mathrm{Tr}(\sigma - rD) = 0, \ |\sigma - rD| \leq 1 & \text{if} \quad |D'| = 0. \end{cases} \tag{2.6.6}$$

If $\sigma = rD + \dfrac{D'}{|D'|}$ with $|D'| \neq 0$, then

$$\mathrm{Tr}(\sigma)\frac{\mathrm{Id}}{N} + \sigma' = r\left(\mathrm{Tr}(D)\frac{\mathrm{Id}}{N} + D'\right) + \frac{D'}{|D'|}, \tag{2.6.7}$$

which is equivalent to

$$\begin{cases} r\,\mathrm{Tr}(D) = \mathrm{Tr}(\sigma), \\ rD' + \dfrac{D'}{|D'|} = \sigma'. \end{cases} \tag{2.6.8}$$

By the second equation $|\sigma'| > 1$, A' and σ' are co-linear, and taking the absolute value we obtain

$$\begin{cases} \mathrm{Tr}(D) = \dfrac{\mathrm{Tr}(\sigma)}{r}, \\ |D'| = \dfrac{|\sigma'| - 1}{r}, \quad D' = \dfrac{|\sigma'| - 1}{r}\dfrac{\sigma'}{|\sigma'|}. \end{cases} \tag{2.6.9}$$

Then

$$D = \frac{\mathrm{Tr}(\sigma)}{rN}\,\mathrm{Id} + \frac{|\sigma'| - 1}{r}\frac{\sigma'}{|\sigma'|}. \tag{2.6.10}$$

Besides, if $\mathrm{Tr}(\sigma - rD) = 0$, $|\sigma - rD| \leq 1$ with $|D'| = 0$, then

$$D = \frac{\mathrm{Tr}(\sigma)}{rN}\,\mathrm{Id}, \ \text{and} \ |\sigma'| \leq 1.$$

Therefore we obtain in any case

$$\left(\mathrm{Id}+\frac{\partial F}{r}\right)^{-1}\left(\frac{\sigma}{r}\right)=\frac{\mathrm{Tr}(\sigma)}{rN}\,\mathrm{Id}+\left(\frac{|\sigma'|-1}{r}\right)_{+}\frac{\sigma'}{|\sigma'|}. \tag{2.6.11}$$

With (2.6.5) we deduce that

$$\mathbb{P}_{\mathcal{A}}(\sigma)=\begin{cases}\sigma' & \text{if } |\sigma'|\leq 1,\\ \dfrac{\sigma'}{|\sigma'|} & \text{if } |\sigma'|>1.\end{cases}$$

Cam-Clay model

We consider the well-known Cam-Clay viscoplastic model. Defining the set $\mathcal{A}$ or the function F is equivalent, and we choose here to first define $\mathcal{A}$. Afterwards we shall deduce F, as (2.6.21). Writing $\sigma=-p\,\mathrm{Id}+\sigma'$ with $\mathrm{Tr}(\sigma')=0$, we define $\mathcal{A}$ in the variables σ' and p as a half-ellipse,

$$\mathcal{A}:=\left\{\sigma\mid 0\leq p\leq p_0,\ \frac{|\sigma'|^2}{\sin^2\delta}\leq p(p_0-p)\right\}=\left\{\sigma\mid 0\leq p\leq p_0,\ \frac{|\sigma'|^2}{\sin^2\delta}+\left(p-\frac{p_0}{2}\right)^2\leq\frac{p_0^2}{4}\right\}, \tag{2.6.12}$$

where $p_0>0$ and $\sin\delta>0$ are given, see Figure 2.2. Then

$$\begin{aligned}F(D)&=\sup_{0\leq p\leq p_0}\ \sup_{|\sigma'|\leq\sin\delta\sqrt{p_0p-p^2}}(-p\,\mathrm{Tr}(D)+\sigma':D')\\&=\sup_{0\leq p\leq p_0}\left(-p\,\mathrm{Tr}(D)+|D'|\sin\delta\sqrt{p_0p-p^2}\right).\end{aligned} \tag{2.6.13}$$

We set $g(p)=-p\,\mathrm{Tr}(D)+|D'|\sin\delta\sqrt{p_0p-p^2}$. Writing that the derivative of g vanishes at the maximum of the concave function g, we obtain

$$\begin{aligned}-\mathrm{Tr}(D)+|D'|\sin\delta\frac{p_0-2p}{2\sqrt{p_0p-p^2}}&=0,\\|D'|\sin\delta(p_0-2p)&=\mathrm{Tr}(D)2\sqrt{p_0p-p^2}.\end{aligned} \tag{2.6.14}$$

We remark that it implies that $\mathrm{Tr}(D)$ has the same sign as p_0-2p, thus

$$\begin{aligned}\mathrm{Tr}(D)>0&\Rightarrow p<\frac{p_0}{2},\\\mathrm{Tr}(D)<0&\Rightarrow p>\frac{p_0}{2}.\end{aligned} \tag{2.6.15}$$

Then we take the square in (2.6.14), giving

$$|D'|^2\sin^2\delta(p_0^2-4p_0p+4p^2)=4\,\mathrm{Tr}(D)^2(p_0p-p^2), \tag{2.6.16}$$

$$p^2-pp_0+\frac{|D'|^2\sin^2\delta}{4(|D'|^2\sin^2\delta+\mathrm{Tr}(D)^2)}p_0^2=0, \tag{2.6.17}$$

$$p=\frac{p_0}{2}\pm\frac{p_0}{2}\frac{\mathrm{Tr}(D)}{\sqrt{|D'|^2\sin^2\delta+\mathrm{Tr}(D)^2}}. \tag{2.6.18}$$

Recalling from (2.6.15) that $\mathrm{Tr}(D)$ has the same sign as p_0-2p, we get that the sign is minus, thus

$$p=\frac{p_0}{2}\left(1-\frac{Y}{\sqrt{1+Y^2}}\right),\ \text{with } Y=\frac{\mathrm{Tr}(D)}{|D'|\sin\delta}. \tag{2.6.19}$$

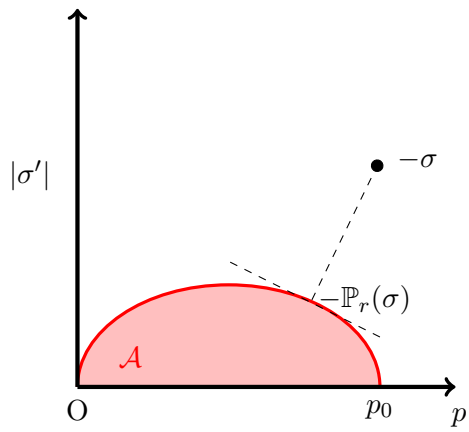


Figure 2.2: Set $\mathcal{A}$ in the Cam-Clay model, and orthogonal projection.

The relation between p and Y can also be inverted as

$$Y = \frac{p_0 - 2p}{2\sqrt{p_0 p - p^2}}. \tag{2.6.20}$$

Hence (2.6.13) yields

$$F(D) = -\frac{p_0}{2}\left(1 - \frac{Y}{\sqrt{1+Y^2}}\right)\mathrm{Tr}(D) + |D'|\sin\delta\frac{p_0}{2\sqrt{1+Y^2}}.$$

Using that $\mathrm{Tr}(D) = Y|D'|\sin\delta$, it can be recast as

$$F(D) = \frac{p_0}{2}\frac{\sin\delta|D'|}{\sqrt{1+Y^2}+Y}.$$

Thus

$$F(D) = \frac{p_0}{2}\frac{\sin\delta|D'|}{Y+\sqrt{1+Y^2}}, \quad \text{with } Y = \frac{\mathrm{Tr}(D)}{|D'|\sin\delta}. \tag{2.6.21}$$

We have to mention that according to Lemma 2.1.3(c), the point σ where the supremum (2.6.13) is attained gives $\partial F(D)$.

For $|D'| = 0$ one has $F(D) = 0$ if $\mathrm{Tr}(D) \geq 0$, $F(D) = -p_0\,\mathrm{Tr}(D)$ if $\mathrm{Tr}(D) \leq 0$.

Finding the orthogonal projection on the half ellipse

Since the orthogonal projection on $\mathcal{A}$ is involved in Theorem 2.5.5, it is useful to compute it. For a stress σ_b on the boundary of $\mathcal{A}$, one has an outer normal n_b, such that for a small variation $\delta\sigma_b$ one has $n_b : \delta\sigma_b = 0$. In order to find the orthogonal projection of $\sigma = -p\,\mathrm{Id} + \sigma'$ on $\mathcal{A}$ ($\sigma \notin \mathcal{A}$), according to Theorem 2.5.5 one has to find $\sigma_b \in \partial\mathcal{A}$ and $\lambda \geq 0$ such that $\sigma = \sigma_b + \lambda n_b$. Then $\mathbb{P}_{\mathcal{A}}(\sigma) = \sigma_b$. One has

$$\sigma_b = -p_b\,\mathrm{Id} + \sigma'_b\sin\delta\sqrt{p_b(p_0 - p_b)}, \tag{2.6.22}$$

with $\mathrm{Tr}(\sigma'_b) = 0$, $|\sigma'_b| = 1$, and $0 \leq p_b \leq p_0$. Assume that $0 < p_b < p_0$. Then for a small variation $\delta\sigma_b$ of σ_b one has

$$\delta\sigma_b = \Big(-\mathrm{Id} + \sigma'_b\sin\delta\frac{p_0 - 2p_b}{2\sqrt{p_b(p_0 - p_b)}}\Big)\delta p_b + \delta\sigma'_b\sin\delta\sqrt{p_b(p_0 - p_b)}. \tag{2.6.23}$$

Then since $\mathrm{Tr}(\delta\sigma'_b) = 0$, $\sigma'_b : \delta\sigma'_b = 0$, we compute

$$\delta\sigma_b : \mathrm{Id} = -N\delta p_b, \qquad \delta\sigma_b : \sigma'_b = \sin\delta\frac{p_0 - 2p_b}{2\sqrt{p_b(p_0 - p_b)}}\delta p_b. \tag{2.6.24}$$

Thus we get

$$\delta\sigma_b : \Big({\sigma'}_b + \frac{\sin\delta}{N}\frac{p_0 - 2p_b}{2\sqrt{p_b(p_0-p_b)}}\,\mathrm{Id}\Big) = 0, \tag{2.6.25}$$

and it follows that we can take

$$n_b = {\sigma'}_b + \frac{\sin\delta}{N}\frac{p_0 - 2p_b}{2\sqrt{p_b(p_0-p_b)}}\,\mathrm{Id}\,. \tag{2.6.26}$$

This formula could also be deduced from the computation of the supremum above, by Proposition 2.5.3(d), with the help of Lemma 2.1.3(b,c), and using the inversion formula (2.6.20). Thus in order to find the orthogonal projection of $\sigma = -p\,\mathrm{Id} + \sigma'$ on $\mathcal{A}$ ($\sigma \notin \mathcal{A}$), one has to find $\sigma_b \in \partial\mathcal{A}$ and $\lambda \geq 0$ such that

$$\sigma \equiv -p\,\mathrm{Id} + \sigma' = -p_b\,\mathrm{Id} + {\sigma'}_b \sin\delta\sqrt{p_b(p_0-p_b)} + \lambda\Big({\sigma'}_b + \frac{\sin\delta}{N}\frac{p_0 - 2p_b}{2\sqrt{p_b(p_0-p_b)}}\,\mathrm{Id}\Big). \tag{2.6.27}$$

Taking the trace and the deviator parts we get that σ' and ${\sigma'}_b$ are aligned (thus ${\sigma'}_b = \sigma'/|\sigma'|$) and the equations

$$-p = -p_b + \lambda\frac{\sin\delta}{N}\frac{p_0 - 2p_b}{2\sqrt{p_b(p_0-p_b)}}, \tag{2.6.28}$$

$$|\sigma'| = \lambda + \sin\delta\sqrt{p_b(p_0-p_b)}. \tag{2.6.29}$$

These two equations determine $0 \leq p_b \leq p_0$ and $\lambda \geq 0$, that should be unique under the assumption that $\sigma \notin \mathcal{A}$ i.e. either $p < 0$, or $p > p_0$, or $0 \leq p \leq p_0$ and $|\sigma'| > \sin\delta\sqrt{p(p_0-p)}$. Inserting the expression of λ from (2.6.29) into (2.6.28) yields

$$-p = -p_b + \frac{\sin\delta}{N}\frac{p_0 - 2p_b}{2\sqrt{p_b(p_0-p_b)}}\Big(|\sigma'| - \sin\delta\sqrt{p_b(p_0-p_b)}\Big), \tag{2.6.30}$$

which is a quartic equation in p_b. This can be rewritten equivalently in terms of Y_b related to p_b by (2.6.19), as

$$\frac{\sin\delta}{N}\frac{2}{p_0}|\sigma'|Y_b + \Big(1 - \frac{\sin^2\delta}{N}\Big)\frac{Y_b}{\sqrt{1+Y_b^2}} + \frac{2p}{p_0} - 1 = 0. \tag{2.6.31}$$

As soon as $|\sigma'| > 0$ this equation always has a single real solution Y_b. One can check that this solution gives $\lambda \geq 0$ provided that $\sigma \notin \mathcal{A}$. The solution Y_b to (2.6.31) can be computed by the Newton method. In the case $|\sigma'| = 0$ we simply have $p_b = 0$ if $p < 0$, $p_b = p_0$ if $p > p_0$. Finally we get $\mathbb{P}_{\mathcal{A}}(\sigma) = \sigma_b = -p_b\,\mathrm{Id} + \dfrac{\sigma'}{|\sigma'|}\sin\delta\sqrt{p_b(p_0-p_b)}$.

Drucker-dilatant model

Next we introduce the Drucker-dilatant viscoplastic model. As for the Cam-Clay model, let us start from defining $\mathcal{A}$. We take it as a triangle, as shown on Figure 2.3,

$$\mathcal{A} := \Big\{\sigma \mid 0 \leq p \leq p_0 \text{ and } |\sigma'| \leq p\sin\psi\Big\}, \tag{2.6.32}$$

where $p_0 > 0$ and $\sin\Psi > 0$ are given. According to Proposition 2.5.1 we can then formulate F as

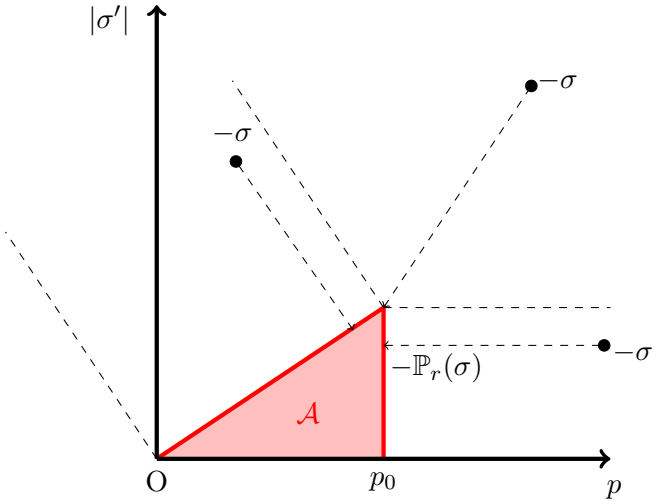


Figure 2.3: Set $\mathcal{A}$ in the Drucker-dilatant model with 4 possible cases of projections.

$$\begin{aligned}
F(D) &= \sup_{\sigma\in\mathcal{A}}(\sigma : D) = \sup_{0\le p\le p_0}\ \sup_{|\sigma'|\le p\sin\psi}(-p\,\mathrm{Tr}(D) + \sigma' : D') \\
&= \sup_{0\le p\le p_0}\big(-p\,\mathrm{Tr}(D) + |D'|p\sin\psi\big) \\
&= \sup_{0\le p\le p_0} p(\sin\psi|D'| - \mathrm{Tr}(D)) \\
&= \begin{cases} 0 & \text{if } \sin\psi|D'| - \mathrm{Tr}(D) \le 0, \\ p_0(\sin\psi|D'| - \mathrm{Tr}(D)) & \text{if } \sin\psi|D'| - \mathrm{Tr}(D) \ge 0. \end{cases}
\end{aligned}$$

Thus F can be rewritten more concisely under the form

$$F(D) = p_0(\sin\psi|D'| - \mathrm{Tr}(D))_+. \tag{2.6.33}$$

Using Lemma 2.1.3(c), we get that for a given D, the stress $\sigma = -p\,\mathrm{Id} + \sigma' \in \partial F(D)$ is characterized by $\sigma' = p\sin\psi\frac{D'}{|D'|}$ if $D' \neq 0$, $|\sigma'| \le p\sin\psi$ if $D' = 0$, and

$$\begin{cases} p = 0 & \text{if } \sin\psi|D'| - \mathrm{Tr}(D) < 0, \\ 0 \le p \le p_0 & \text{if } \sin\psi|D'| - \mathrm{Tr}(D) = 0, \\ p = p_0 & \text{if } \sin\psi|D'| - \mathrm{Tr}(D) > 0. \end{cases} \tag{2.6.34}$$

In particular, having a pressure $p < p_0$ implies that $\mathrm{Tr}(D) \ge 0$. This justifies the name "Drucker-dilatant": as long as the pressure does not take the maximal value, the material dilates. Moreover, as long as $0 < p < p_0$, we must have the dilatancy law $\mathrm{Tr}(D) = \sin\psi|D'|$.

Finding the orthogonal projection on the triangle

We take $\sigma = -p\,\mathrm{Id} + \sigma' \notin \mathcal{A}$, and we have to find $\mathbb{P}_{\mathcal{A}}(\sigma) = \sigma_b \in \partial\mathcal{A}$, such that $\sigma = \sigma_b + \lambda n_b$, with n_b the external normal to $\partial\mathcal{A}$ at σ_b, and $\lambda \ge 0$.

Let us first consider the case of the inclined boundary. Then

$$\sigma_b = -p_b\,\mathrm{Id} + p_b\sin\psi\sigma'_b, \tag{2.6.35}$$

with $\mathrm{Tr}(\sigma'_b) = 0$, $|\sigma'_b| = 1$. Then $\delta\sigma_b = -\mathrm{Id}\,\delta p_b + \sin\psi\sigma'_b\delta p_b + p_b\sin\psi\delta\sigma'_b$. We have then $\mathrm{Tr}(\delta\sigma_b) = -N\delta p_b$, $\delta\sigma_b : \sigma'_b = \sin\psi\delta p_b$. Thus $\delta\sigma_b : (\sigma'_b + \mathrm{Id}\sin\psi/N) = 0$. Therefore we can take

$n_b = \sigma'_b + \mathrm{Id}\sin\psi/N$. We write that $\sigma = \sigma_b + \lambda n_b$, which gives

$$-p = -p_b + \lambda\frac{\sin\psi}{N}, \qquad \sigma' = (\lambda + p_b\sin\psi)\sigma'_b. \tag{2.6.36}$$

If follows that

$$\sigma'_b = \frac{\sigma'}{|\sigma'|}, \quad |\sigma'| = \lambda + p_b\sin\psi, \quad -p = -p_b + \lambda\frac{\sin\psi}{N}. \tag{2.6.37}$$

Plugging the value of λ into the equation on p we obtain

$$-p = -p_b + (|\sigma'| - p_b\sin\psi)\frac{\sin\psi}{N}, \tag{2.6.38}$$

thus

$$p_b = \frac{p + |\sigma'|\frac{\sin\psi}{N}}{1 + \frac{\sin^2\psi}{N}}, \qquad \sigma_b = \frac{p + |\sigma'|\frac{\sin\psi}{N}}{1 + \frac{\sin^2\psi}{N}}\Big(-\mathrm{Id} + \sin\psi\frac{\sigma'}{|\sigma'|}\Big). \tag{2.6.39}$$

This formula is valid if $0 \le p_b \le p_0$, and $\lambda \ge 0$, which gives the conditions

$$0 \le p + |\sigma'|\frac{\sin\psi}{N} \le \Big(1 + \frac{\sin^2\psi}{N}\Big)p_0, \quad \text{and} \quad p\sin\psi \le |\sigma'|. \tag{2.6.40}$$

Next let us consider the case of the right boundary. In this case $\sigma_b = -p_0\,\mathrm{Id} + \mu\sigma'_b$, with $\mathrm{Tr}(\sigma'_b) = 0$, $|\sigma'_b| = 1$, $0 \le \mu \le p_0\sin\psi$. Then $\delta\sigma_b = \delta\mu\sigma'_b + \mu\delta\sigma'_b$, $\mathrm{Tr}(\delta\sigma_b) = 0$. It follows that we can take $n_b = -\mathrm{Id}$. Then writing that $\sigma = -p\,\mathrm{Id} + \sigma' = \sigma_b + \lambda n_b$ we obtain

$$-p = -p_0 - \lambda, \qquad \sigma' = \mu\sigma'_b. \tag{2.6.41}$$

It follows that

$$\sigma'_b = \frac{\sigma'}{|\sigma'|}, \quad |\sigma'| = \mu. \tag{2.6.42}$$

Finally $\sigma_b = -p_0\,\mathrm{Id} + \sigma'$. This is valid when $p \ge p_0$ and $|\sigma'| \le p_0\sin\psi$.

For the other values of σ, σ_b is one of the corners. Thus finally, for $\sigma \notin \mathcal{A}$,

$$\mathbb{P}_{\mathcal{A}}(\sigma) = \begin{cases} 0 & \text{if } p + \frac{\sin\psi}{N}|\sigma'| \le 0, \\ \dfrac{p + \frac{\sin\psi}{N}|\sigma'|}{1 + \frac{\sin^2\psi}{N}}\Big(-\mathrm{Id} + \sin\psi\frac{\sigma'}{|\sigma'|}\Big) & \text{if } 0 \le p + \frac{\sin\psi}{N}|\sigma'| \le \big(1 + \frac{\sin^2\psi}{N}\big)p_0, p\sin\psi \le |\sigma'|, \\ -p_0\,\mathrm{Id} + p_0\sin\psi\frac{\sigma'}{|\sigma'|} & \text{if } p + \frac{\sin\psi}{N}|\sigma'| \ge (1 + \frac{\sin^2\psi}{N})p_0, |\sigma'| \ge p_0\sin\psi, \\ -p_0\,\mathrm{Id} + \sigma' & \text{if } |\sigma'| \le p_0\sin\psi,\ p \ge p_0. \end{cases} \tag{2.6.43}$$

A logarithmic model

In this model we define $\mathcal{A}$ as

$$\mathcal{A} := \Big\{\sigma \mid 0 \le p \le ep_0 \text{ and } |\sigma'| \le \lambda p\left(1 - \ln\frac{p}{p_0}\right)\Big\}, \tag{2.6.44}$$

for some given $p_0 > 0$, $\lambda > 0$, and where $e = \exp(1)$, see Figure 2.4. Using Proposition 2.5.1 we

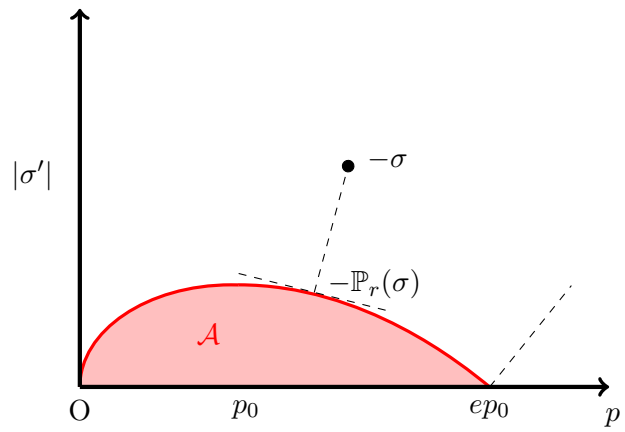


Figure 2.4: Set $\mathcal{A}$ in the logarithmic model, and orthogonal projection.

compute F as

$$\begin{aligned} F(D) &= \sup_{\sigma\in\mathcal{A}}(\sigma : D) \\ &= \sup_{0\le p\le ep_0}\ \sup_{|\sigma'|\le\lambda p(1-\ln(p/p_0))}(-p\operatorname{Tr}(D)+\sigma' : D') \\ &= \sup_{0\le p\le ep_0}\left(-p\operatorname{Tr}(D)+|D'|\lambda p\left(1-\ln\frac{p}{p_0}\right)\right) \\ &= \lambda\sup_{0\le p\le ep_0}\left(p\left(|D'|-\frac{\operatorname{Tr}(D)}{\lambda}\right)-|D'|p\ln\frac{p}{p_0}\right). \end{aligned}$$

In the case $|D'|=0$, one has

$$F(D)=\begin{cases}0 & \text{if } \operatorname{Tr}(D)\ge 0,\\ -ep_0\operatorname{Tr}(D) & \text{if } \operatorname{Tr}(D)\le 0.\end{cases} \tag{2.6.45}$$

Otherwise if $|D'|>0$ we set $g(p)=p\left(|D'|-\frac{\operatorname{Tr}(D)}{\lambda}\right)-|D'|p\ln\frac{p}{p_0}$. The concave function g attains its maximum at the point p_{max} where the derivative of g vanishes. In other words p_{max} solves

$$-\frac{\operatorname{Tr}(D)}{\lambda}-|D'|\ln\frac{p_{max}}{p_0}=0,$$

thus

$$p_{max}=p_0\exp\left(-\frac{\operatorname{Tr}(D)}{\lambda|D'|}\right). \tag{2.6.46}$$

Knowing that we have the constrain that $0\le p\le ep_0$, we deduce that

- If $p_{max}\le ep_0$ then the supremum is attained at $p=p_{max}$.
- If $p_{max}>ep_0$ then the supremum is attained at $p=ep_0$.

We check that

$$p_{max}\le ep_0 \iff p_0\exp\left(\frac{-\operatorname{Tr}(D)}{\lambda|D'|}\right)\le ep_0 \iff \frac{-\operatorname{Tr}(D)}{\lambda|D'|}\le 1 \iff \operatorname{Tr}(D)\ge-\lambda|D'|.$$

It follows that

- If $\mathrm{Tr}(D) \geq -\lambda|D'|$ then the supremum is attained at $p = p_{max}$.
- If $\mathrm{Tr}(D) < -\lambda|D'|$ then the supremum is attained at $p = ep_0$.

Then for $\mathrm{Tr}(D) \geq -\lambda|D'|$, taking $p = p_{max}$ gives

$$\begin{aligned} F(D) &= \lambda\left(p_{max}\left(|D'| - \frac{\mathrm{Tr}(D)}{\lambda}\right) - |D'|p_{max}\ln\frac{p_{max}}{p_0}\right) \\ &= \lambda\left(p_0\exp\left(-\frac{\mathrm{Tr}(D)}{\lambda|D'|}\right)\left(|D'| - \frac{\mathrm{Tr}(D)}{\lambda}\right) + |D'|p_0\exp\left(-\frac{\mathrm{Tr}(D)}{\lambda|D'|}\right)\frac{\mathrm{Tr}(D)}{\lambda|D'|}\right) \\ &= \lambda p_0|D'|\exp\left(-\frac{\mathrm{Tr}(D)}{\lambda|D'|}\right). \end{aligned}$$

Otherwise if $\mathrm{Tr}(D) < -\lambda|D'|$, taking $p = ep_0$ gives

$$F(D) = \lambda\left(ep_0\left(|D'| - \frac{\mathrm{Tr}(D)}{\lambda}\right) - |D'|ep_0\right) = -ep_0\,\mathrm{Tr}(D). \tag{2.6.47}$$

Finally we obtain the general formula

$$F(D) = \begin{cases} \lambda p_0|D'|\exp\left(-\dfrac{\mathrm{Tr}(D)}{\lambda|D'|}\right) & \text{if } \mathrm{Tr}(D) \geq -\lambda|D'|, \\ -ep_0\,\mathrm{Tr}(D) & \text{if } \mathrm{Tr}(D) \leq -\lambda|D'|. \end{cases} \tag{2.6.48}$$

Moreover, using Lemma 2.1.3 we obtain that for $D' \neq 0$, $\partial F(D)$ is a single point σ obtained by $\sigma' = \lambda p(1 - \ln\frac{p}{p_0})\frac{D'}{|D'|}$ and

$$p = \begin{cases} p_0\exp\left(-\dfrac{\mathrm{Tr}(D)}{\lambda|D'|}\right) & \text{if } \mathrm{Tr}(D) \geq -\lambda|D'|, \\ ep_0 & \text{if } \mathrm{Tr}(D) \leq -\lambda|D'|. \end{cases} \tag{2.6.49}$$

In the case $D' = 0$ and $\mathrm{Tr}(D) \neq 0$, $\partial F(D)$ is a single point σ defined as $\sigma = 0$ if $\mathrm{Tr}(D) > 0$, $\sigma = -ep_0\,\mathrm{Id}$ if $\mathrm{Tr}(D) < 0$.

We observe in particular that for $D' \neq 0$ and $\mathrm{Tr}(D) \geq -\lambda|D'|$, we have

$$\sigma' = p\Big(\lambda + \frac{\mathrm{Tr}(D)}{|D'|}\Big)\frac{D'}{|D'|}, \tag{2.6.50}$$

which is a friction law of the type expected in granular materials: proportional to the pressure with a coefficient which is the sum of a constant and the dilatancy ratio $\mathrm{Tr}(D)/|D'|$. The dilatancy law is

$$\mathrm{Tr}(D) = -\lambda|D'|\ln\frac{p}{p_0}, \tag{2.6.51}$$

with a critical pressure p_0.

Finding the orthogonal projection on $\mathcal{A}$ for the logarithmic model

We take $\sigma = -p\,\mathrm{Id} + \sigma' \notin \mathcal{A}$, and we have to find $\mathbb{P}_{\mathcal{A}}(\sigma) = \sigma_b \in \partial\mathcal{A}$, such that $\sigma = \sigma_b + \gamma n_b$, with n_b the external normal to $\partial\mathcal{A}$ at σ_b and $\gamma \geq 0$. One has

$$\sigma_b = -p_b\,\mathrm{Id} + \lambda p_b\left(1 - \ln\frac{p_b}{p_0}\right)\sigma'_b, \tag{2.6.52}$$

with $\mathrm{Tr}(\sigma'_b) = 0$, $|\sigma'_b| = 1$. Then

$$\delta\sigma_b = -\,\mathrm{Id}\,\delta p_b - \lambda \ln \frac{p_b}{p_0}\sigma'_b \delta p_b + \lambda p_b \left(1 - \ln \frac{p_b}{p_0}\right) \delta\sigma'_b. \tag{2.6.53}$$

Then since $\mathrm{Tr}(\delta\sigma'_b) = 0$, $\sigma'_b : \delta\sigma'_b = 0$, we compute

$$\delta\sigma_b : \mathrm{Id} = -N\delta p_b, \qquad \delta\sigma_b : \sigma'_b = -\lambda \ln \frac{p_b}{p_0}\delta p_b. \tag{2.6.54}$$

Thus we get $\delta\sigma_b : \left(\sigma'_b - \frac{\lambda}{N} \ln \frac{p_b}{p_0}\,\mathrm{Id}\right) = 0$, and it follows that $n_b = \sigma'_b - \frac{\lambda}{N} \ln \frac{p_b}{p_0}\,\mathrm{Id}$. Writing that $\sigma = \sigma_b + \gamma n_b$ we obtain

$$-\,p = -p_b - \gamma \frac{\lambda}{N} \ln \frac{p_b}{p_0}, \qquad \sigma' = \left(\gamma + \lambda p_b \Big(1 - \ln \frac{p_b}{p_0}\Big)\right) \sigma'_b. \tag{2.6.55}$$

If follows that $\sigma'_b = \dfrac{\sigma'}{|\sigma'|}$,

$$|\sigma'| = \gamma + \lambda p_b \Big(1 - \ln \frac{p_b}{p_0}\Big), \tag{2.6.56}$$

$$p = p_b + \gamma \frac{\lambda}{N} \ln \frac{p_b}{p_0}. \tag{2.6.57}$$

Inserting the expression of γ from (2.6.56) into (2.6.57) yields

$$p = p_b + \frac{\lambda}{N} \ln \frac{p_b}{p_0} \left(|\sigma'| - \lambda p_b \Big(1 - \ln \frac{p_b}{p_0}\Big)\right). \tag{2.6.58}$$

We observe graphically on Figure 2.4 that for $\sigma = -p\,\mathrm{Id} + \sigma' \notin \mathcal{A}$, the projection $\mathbb{P}_{\mathcal{A}}(\sigma) = \sigma_b$ is given by

- If $|\sigma'| = 0$ and $p \leq 0$ then $\sigma_b = 0$,
- If $|\sigma'| \leq \dfrac{N}{\lambda}(p - ep_0)$ then $\sigma_b = -ep_0\,\mathrm{Id}$,
- If $|\sigma'| > 0$ and $|\sigma'| > \dfrac{N}{\lambda}(p - ep_0)$ then σ_b is given by (2.6.52) with $\sigma'_b = \dfrac{\sigma'}{|\sigma'|}$ and $p_b \in (0, ep_0)$ solution to (2.6.58) such that $p_b\big(1 - \ln \frac{p_b}{p_0}\big) \leq \dfrac{|\sigma'|}{\lambda}$.

Remark: In the three last models, the assumption that F is convex, lower semi-continuous, proper is not obvious from its expression. Nevertheless it follows by construction from Proposition 2.5.1 that depicts the relation between the function F and the set $\mathcal{A}$, knowing that $\mathcal{A}$ is convex, closed and non-empty, which is true when it is defined by an inequality of the form $|\sigma'| \leq f(p)$ with f concave.

We remark that in the Bingham model, when $\sigma \in \mathcal{A}$ the pressure $p = -\,\mathrm{Tr}(\sigma)/N$ can take negative values. The other models presented above verify $p \geq 0$ when $\sigma \in \mathcal{A}$, which is more physical. One could also ask the stronger condition $\sigma \leq 0$ in the sense of symmetric matrices. A sufficient condition for that is $|\sigma'| \leq \sqrt{\frac{N}{N-1}}p$ (it is also necessary if $N = 2$). It indicates that the set $\mathcal{A}$ should be cut below this line, similarly as in the Drucker-dilatant model. In this case, and if $0 \in \mathcal{A}$, we see geometrically that $|D'| < \sqrt{\frac{N-1}{N}}\,\mathrm{Tr}(D) \Rightarrow \partial F(D) = \{0\}$. A related property is that $|\,\mathrm{Tr}(D)| < \sqrt{\frac{N}{N-1}}|D'|$ implies that the eigenvalues of D are not all of the same sign.

Chapter 3

Existence of the solution for the steady viscoplastic model with nonlinear rheology

Before going to the numerical approximation of (2.0.1), an existence theory is necessary. Moreover the analysis of Chapter 4 shows that we need a formulation in the local sense (not only in integral form). This is sometimes called the Euler-Lagrange equation, and it does not follow directly from the variational theory.

In the case of the (possibly viscous) incompressible Bingham model, i.e. $F(D) = \frac{\eta}{2}|D|^2 + \sigma_0|D|$, one can find in the literature some discussions on the existence of solutions for the inviscid case $\eta = 0$, as in [39] or [14]. In all the classical results of [26] the analysis relies on the viscous term, and strongly depends on the assumption $\eta > 0$. Meanwhile in [14] the results are established for the case $\eta = 0$. These results show that the two cases (viscous or inviscid) have to be treated by a slighly different analysis, relying on the theory of variational inequalities and monotone operators [16], but with a different choice of Hilbert space.

In this chapter we consider the viscoplastic model (2.0.1) with $\sigma \in \partial F(Du)$ where F is convex, proper, l.s.c. on $\mathbb{M}^s_{N\times N}(\mathbb{R})$ the space of symmetric square matrices of size N, and satisfies some growth conditions. We recall that the problem can be formulated as

$$\inf_{u\in H} \int_\Omega \left(\alpha \frac{|u|^2}{2} + F(Du) - f \cdot u \right) \mathrm{dx}, \tag{3.0.1}$$

where H is an appropriate Hilbert space of functions $u(x)$ with values in $\mathbb{R}^N$. The key issue is to define the space H for a general nonlinearity F, in such a way that $u \mapsto \int F(Du)$ is convex, proper and l.s.c. on H. Then one has to prove also the Euler-Lagrange equations, i.e. the local formulation (2.0.1).

3.1 With viscosity

We first consider the case with additional viscosity $\eta > 0$, i.e.

$$\inf_{u\in H} \int_\Omega \left(\alpha \frac{|u|^2}{2} + \frac{\eta}{2}|Du|^2 + F(Du) - f \cdot u \right) \mathrm{dx}\,. \tag{3.1.1}$$

Another way to formulate it would be to consider the problem (3.0.1) with a coercive nonlinearity $F(D) + \frac{\eta}{2}|D|^2$. We recall that Hypothesis 1 means that F is convex, l.s.c. and proper. This will always be assumed. Then from Section 2.1 we have that F is necessarily bounded below by an affine function. With the notation of Proposition 2.1.4 we have the following result.

Lemma 3.1.1. *Suppose that F satisfies Hypothesis 1. Then $|\operatorname{prox}_\varepsilon F(0)|$ is bounded independently of ε as long as $0 < \varepsilon \leq \varepsilon_0$.*

Proof. Since F is proper, there exists D_0 such that $F(D_0) < \infty$. Then since $\operatorname{prox}_\varepsilon F$ is $1-$Lipschitz, one has $|\operatorname{prox}_\varepsilon F(0) - \operatorname{prox}_\varepsilon F(D_0)| \leq |D_0|$. Thus proving that $\operatorname{prox}_\varepsilon F(0)$ is bounded independently of ε is equivalent to proving that $\operatorname{prox}_\varepsilon F(D_0)$ is bounded independently of ε. The definition of the Moreau envelope gives

$$F_\varepsilon(D_0) = \inf_{\overline{D}} \left\{ F(\overline{D}) + \frac{|\overline{D} - D_0|^2}{2\varepsilon} \right\} \leq F(D_0). \tag{3.1.2}$$

We have that F is lower bounded by an affine function, thus $F(\overline{D}) \geq A : \overline{D} - B$ for some matrix A and some real number B. It follows that

$$\begin{aligned} F(\overline{D}) + \frac{|\overline{D} - D_0|^2}{2\varepsilon} &\geq A : (\overline{D} - D_0) + A : D_0 - B + \frac{|\overline{D} - D_0|^2}{2\varepsilon} \\ &\geq -\frac{|\overline{D} - D_0|^2}{4\varepsilon} - \varepsilon|A|^2 + A : D_0 - B + \frac{|\overline{D} - D_0|^2}{2\varepsilon} \\ &\geq \frac{|\overline{D} - D_0|^2}{4\varepsilon} + A : D_0 - B - \varepsilon_0|A|^2. \end{aligned} \tag{3.1.3}$$

Therefore $F_\varepsilon(D_0) \geq A : D_0 - B - \varepsilon_0|A|^2$. Moreover there is some $M > 0$ such that the above quantity is larger than $F(D_0)$ for $\frac{|\overline{D} - D_0|}{\sqrt{\varepsilon}} \geq M$. It follows that $F(\overline{D}) + \frac{|\overline{D} - D_0|^2}{2\varepsilon}$ has to attain its infimum inside the ball $|\overline{D} - D_0| \leq M\sqrt{\varepsilon}$. Since $\varepsilon \leq \varepsilon_0$ this set is bounded by $|D_0| + M\sqrt{\varepsilon_0}$. Consequently, $|\operatorname{prox}_\varepsilon F(D_0)| \leq |D_0| + M\sqrt{\varepsilon_0}$ for all $0 < \varepsilon \leq \varepsilon_0$. □

In the following we shall consider several growth assumptions.

Hypothesis 2. *For any $D \in \mathbb{M}^s_{N\times N}(\mathbb{R})$ and $\sigma \in \partial F(D)$ one has*

$$|\sigma| \leq C(1 + |D|),$$

where C is a positive constant independent of D and σ.

Lemma 3.1.2. *Under Hypothesis 1, 2, we have*
(a) the derivative of F_ε verifies

$$|F'_\varepsilon(D)| \leq C'(1 + |D|), \tag{3.1.4}$$

where C' is a positive constant independent of D and $\varepsilon \leq \varepsilon_0$,
(b) F is finite everywhere and for all D

$$|F_\varepsilon(D)| \leq C''(1 + |D|^2), \qquad |F(D)| \leq C''(1 + |D|^2), \tag{3.1.5}$$

where C'' is a constant independent of D and $\varepsilon \leq \varepsilon_0$.

Proof. (a) According to Proposition 2.1.4(e) we have $F'_\varepsilon(D) \in \partial F(\operatorname{prox}_\varepsilon F(D))$, thus

$$|F'_\varepsilon(D)| \leq C(1 + |\operatorname{prox}_\varepsilon F(D)|).$$

Besides, Proposition 2.1.4(c) ensures that $\operatorname{prox}_\varepsilon F$ is 1-Lipschitz, giving

$$|F'_\varepsilon(D)| \leq C(1 + |\operatorname{prox}_\varepsilon F(0)| + |D|). \tag{3.1.6}$$

With Lemma 3.1.1 we obtain (3.1.4).
(b) The inequality (3.1.4) implies that

$$|F_\varepsilon(D) - F_\varepsilon(0)| \leq C'|D|(1 + |D|). \tag{3.1.7}$$

Following (3.1.3), F_ε is lower bounded by an affine function independent of $\varepsilon \leq \varepsilon_0$. Thus $F_\varepsilon(0)$ is bounded below by a constant independent of ε. Moreover according to Proposition 2.1.4(f) we have $F_\varepsilon(0) \uparrow F(0)$ as $\varepsilon \to 0$. If $F_\varepsilon(0)$ were unbounded, (3.1.7) would imply that $F_\varepsilon(D) \to \infty$ for all D, which is not the case since F is proper. Thus $F(0) < \infty$, $|F_\varepsilon(0)|$ remains bounded, and (3.1.7) gives $|F_\varepsilon(D)| \leq C''(1 + |D|^2)$. Since $F_\varepsilon(D) \uparrow F(D)$ we deduce that $F(D) < \infty$ with the same bound. □

Lemma 3.1.3. *Let F satisfy Hypothesis 1 and assume that F is finite everywhere and there exists a constant $C'' > 0$ such that for all D*

$$|F(D)| \leq C''(1 + |D|^2). \tag{3.1.8}$$

Then
(a) For all σ one has

$$F^*(\sigma) \geq \frac{1}{4C''}|\sigma|^2 - C''. \tag{3.1.9}$$

(b) For any D and σ such that $\sigma \in \partial F(D)$ one has

$$|\sigma| \leq 8C''(1 + |D|), \tag{3.1.10}$$

so that F satisfies Hypothesis 2.

Proof. (a) Using the definition of F^* and (3.1.8), one has

$$F^*(\sigma) \geq \sup_D \Big(\sigma : D - C''(1 + |D|^2)\Big). \tag{3.1.11}$$

The supremum is attained at the value of D given by $D = \sigma/(2C'')$. We deduce that

$$F^*(\sigma) \geq \frac{|\sigma|^2}{2C''} - C''\Big(1 + \frac{|\sigma|^2}{4C''^2}\Big), \tag{3.1.12}$$

which yields (3.1.9).
(b) Consider D_1, D_2, and a number $R > 0$ such that $D_1 \neq D_2$ and $|D_1| \leq R$, $|D_2| \leq R$. Let us then set

$$D_3 = D_2 + R\frac{D_2 - D_1}{|D_2 - D_1|}. \tag{3.1.13}$$

We have then $|D_3| \leq 2R$ and

$$D_2 = \frac{R}{|D_2 - D_1| + R}D_1 + \frac{|D_2 - D_1|}{|D_2 - D_1| + R}D_3, \tag{3.1.14}$$

so that by convexity of F,

$$F(D_2) \leq \frac{R}{|D_2 - D_1| + R} F(D_1) + \frac{|D_2 - D_1|}{|D_2 - D_1| + R} F(D_3). \tag{3.1.15}$$

We deduce with (3.1.8) that

$$F(D_2) - F(D_1) \leq \frac{|D_2 - D_1|}{|D_2 - D_1| + R}(F(D_3) - F(D_1)) \leq \frac{|D_2 - D_1|}{R} 2M, \tag{3.1.16}$$

With $M = C''(1 + 4R^2)$. Therefore

$$F(D_2) - F(D_1) \leq |D_2 - D_1| 2C'' \Big(\frac{1}{R} + 4R \Big). \tag{3.1.17}$$

The case $D_2 = D_1$ being trivial, we have now that this inequality holds for all D_1, D_2 such that $|D_1| \leq R$ and $|D_2| \leq R$ Now, given D_1 and D_2, if $\max(|D_1|, |D_2|) \leq 1/2$ then we can take $R = 1/2$, so that $1/R + 4R = 4$. Otherwise we must have $R \geq 1/2$, which implies that $1/R + 4R \leq 2 + 4R$. Thus in any case we get for all D_1, D_2

$$F(D_2) - F(D_1) \leq |D_2 - D_1| 8C'' \Big(1 + \max(|D_1|, |D_2|) \Big). \tag{3.1.18}$$

Consider now D_1 and σ given, such that $\sigma \in \partial F(D_1)$. It follows that for all D_2,

$$F(D_2) \geq F(D_1) + \sigma : (D_2 - D_1). \tag{3.1.19}$$

With (3.1.18) we deduce

$$\sigma : (D_2 - D_1) \leq |D_2 - D_1| 8C'' \Big(1 + \max(|D_1|, |D_2|) \Big). \tag{3.1.20}$$

Taking D_2 in a neighbourhood of D_1 but such that $D_2 - D_1$ takes all the possible directions, we conclude that $|\sigma| \leq 8C''(1 + |D_1|)$, which concludes (3.1.10). □

Remark: Lemmas 3.1.2 and 3.1.3 prove that for a nonlinearity F satisfying Hypothesis 1, one has equivalence between Hypothesis 2 and (3.1.8).

We consider now an open bounded subset Ω of $\mathbb{R}^N$. We shall need a variant of the Sobolev space $H^1(\Omega)$, adapted to the symmetric derivative $Du = (\nabla u + (\nabla u)^t)/2$ where u takes values in $\mathbb{R}^N$. We define

$$H^{1s}(\Omega) = \left\{ u \in L^2(\Omega), \text{ with values in } \mathbb{R}^N, \quad \text{such that } Du \in L^2(\Omega) \right\}, \tag{3.1.21}$$

with the scalar product $\langle u, v \rangle_{L^2} + \langle Du, Dv \rangle_{L^2}$. Then $H^{1s}(\Omega)$ is a Hilbert space. We shall denote by $H_0^{1s}(\Omega)$ the closure of $C_c^\infty(\Omega)$ in $H^{1s}(\Omega)$.

Definition 3.1.4. *Suppose that F satisfies Hypothesis 1, 2 and define for Ω an open bounded subset of $\mathbb{R}^N$, $\psi : H_0^{1s}(\Omega) \to \mathbb{R}$ as*

$$\forall u \in H_0^{1s}(\Omega), \qquad \psi(u) = \int_\Omega F(Du). \tag{3.1.22}$$

Remark: The shorthand notation $\int_\Omega F(Du)$ is used, meaning $\int_\Omega F(Du(x))\,\mathrm{dx}$. According to Lemma 3.1.2(b) and since Ω is bounded, the integral is always finite.

Lemma 3.1.5. *Assume that F satisfies Hypothesis 1, 2. Then ψ is finite everywhere, convex and lower semi-continuous on $H_0^{1s}(\Omega)$.*

Proof. (a) ψ is convex.
If $u, v \in H_0^{1s}(\Omega)$ and $\theta \in (0,1)$, then since F is convex

$$\psi(\theta u + (1-\theta)v) = \int_\Omega F(\theta Du + (1-\theta)Dv) \leq \int_\Omega \big(\theta F(Du) + (1-\theta)F(Dv)\big) = \theta\psi(u) + (1-\theta)\psi(v).$$

(b) ψ is lower semi-continuous.
Suppose $u_n \to u$ in H_0^{1s}. Then one has $Du_n \to Du$ in L^2. Thus after extraction of a subsequence, $Du_n \to Du$ almost everywhere. Since F is l.s.c, one has $F(Du) \leq \underline{\lim} F(Du_n)$ almost everywhere. According to Lemma 3.1.2(b), $F(Du_n)$ is bounded in L^1. Taking the integral over Ω, we get

$$\int_\Omega F(Du) \leq \int_\Omega \underline{\lim} F(Du_n). \tag{3.1.23}$$

Next, $F(D)$ is lower bounded by an affine function $\varphi(D)$. Appying Fatou's Lemma to $F(Du_n) - \varphi(Du_n) \geq 0$, one has

$$\int_\Omega \underline{\lim}(F(Du_n) - \varphi(Du_n)) \leq \underline{\lim} \int_\Omega (F(Du_n) - \varphi(Du_n)).$$

Since $\varphi(Du_n) \to \varphi(Du)$ in L^2 and a.e., we deduce that

$$\int_\Omega \underline{\lim} F(Du_n) \leq \underline{\lim} \int_\Omega F(Du_n). \tag{3.1.24}$$

With (3.1.23) we conclude that $\int F(Du) \leq \underline{\lim} \int F(Du_n)$, i.e. $\psi(u) \leq \underline{\lim}\, \psi(u_n)$. Thus ψ is l.s.c. $\square$

Lemma 3.1.6. *Suppose G is a bounded linear form on $H_0^{1s}(\Omega)$, $u \in H_0^{1s}(\Omega)$, ψ is a convex, proper, l.s.c. function on $H_0^{1s}(\Omega)$, η, Δt are positive constants. Then the two following sets of inequalities are equivalent*

$$\forall v \in H_0^{1s}(\Omega) \quad \frac{1}{\Delta t}\int_\Omega u\cdot(v-u) + \frac{\eta}{2}\int_\Omega |Dv|^2 - \frac{\eta}{2}\int_\Omega |Du|^2 + \psi(v) \geq \psi(u) + G(v-u), \tag{3.1.25}$$

$$\forall v \in H_0^{1s}(\Omega) \quad \frac{1}{\Delta t}\int_\Omega u\cdot(v-u) + \eta\int_\Omega Du : (Dv - Du) + \psi(v) \geq \psi(u) + G(v-u). \tag{3.1.26}$$

Proof. Since $g(D) = |D|^2/2$ for all $D \in M^s_{N\times N}(\mathbb{R})$ is a convex function and $g'(D) = D$, we have

$$\frac{|Dv|^2}{2} \geq \frac{|Du|^2}{2} + Du : (Dv - Du).$$

This gives directly that (3.1.26) implies (3.1.25).

Reciprocally, assume that (3.1.25) hold. Take $v = u + \theta(v^0 - u)$ where $\theta \in [0,1]$ and $v^0 \in H_0^{1s}(\Omega)$. Then $Dv = Du + \theta(Dv^0 - Du)$. Taking the square of both sides and the integral over Ω we get

$$\int_\Omega |Dv|^2 - \int_\Omega |Du|^2 = 2\theta \int_\Omega Du : (Dv^0 - Du) + \theta^2 \int_\Omega |Dv^0 - Du|^2. \tag{3.1.27}$$

Multiplying by $\frac{\eta}{2}$, substituting in (3.1.25) and using that by convexity of ψ one has $\psi\left(u + \theta(v^0 - u)\right) \leq (1-\theta)\psi(u) + \theta\psi(v^0)$ (and observing that (3.1.25) implies that $\psi(u) < \infty$), we obtain

$$\frac{\theta}{\Delta t}\int_\Omega u \cdot (v^0 - u) + \theta\eta \int_\Omega Du : (Dv^0 - Du) + \frac{\eta}{2}\theta^2 \int_\Omega |Dv^0 - Du|^2 + \theta\psi(v^0) - \theta\psi(u) \geq \theta G(v^0 - u).$$

Dividing by $\theta > 0$ and letting $\theta \to 0$, we recover (3.1.26). □

Remark: (a) Since $\int_\Omega Du : (Dv - Du) = \frac{1}{2}\left(\int_\Omega |Dv|^2 - \int_\Omega |Du|^2 - \int_\Omega |Dv - Du|^2\right)$, the variational inequality (3.1.26) can be written as

$$\forall v \in H_0^{1s}(\Omega) \quad \frac{1}{\Delta t}\int_\Omega u\cdot(v-u) + \frac{\eta}{2}\int_\Omega |Dv|^2 - \frac{\eta}{2}\int_\Omega |Du|^2 - \frac{\eta}{2}\int_\Omega |Dv - Du|^2 + \psi(v) \geq \psi(u) + G(v-u).$$

This inequality looks stronger than (3.1.25). However, Lemma 3.1.6 says that there is indeed equivalence with (3.1.25).
(b) The variational formulation (3.1.26) is powerful and plays a crucial role to establish estimates, see [26] and references therein. However it can be difficult to establish it when numerical approximations are involved. Indeed when using test functions in finite elements spaces, a large number of inequalities arise. Then it is useful to use the Euler-Lagrange local formulation, see Theorem 3.1.9 below.

Proposition 3.1.7. *Assume that F satisfies Hypothesis 1, 2. Then there is a unique solution $u \in H_0^{1s}(\Omega)$ to*

$$\forall v \in H_0^{1s}(\Omega) \quad \frac{1}{\Delta t}\int_\Omega u\cdot(v-u) + \int_\Omega F(Dv) - \int_\Omega F(Du) + \eta\int_\Omega Du : (Dv - Du) \geq G(v-u), \tag{3.1.28}$$

where G is a bounded linear form on $H_0^{1s}(\Omega)$, and η, Δt are positive constants (the correspondence with (3.1.1) is $\alpha = 1/\Delta t$).

Proof. With ψ coming from Definition 3.1.4, the set of inequalities (3.1.28) identifies to (3.1.26). Thus according to Lemma 3.1.6 it is equivalent to (3.1.25). Using the argument of the proof of Lemma 3.1.6 applied to the first term $\int u \cdot (v - u)$ we obtain that it is also equivalent to

$$\forall v \in H_0^{1s}(\Omega) \quad \frac{1}{\Delta t}\int_\Omega \frac{1}{2}(|v|^2 - |u|^2) + \frac{\eta}{2}\int_\Omega |Dv|^2 - \frac{\eta}{2}\int_\Omega |Du|^2 + \psi(v) - \psi(u) \geq G(v-u),$$

which means that u is a minimum over $H_0^{1s}(\Omega)$ of the functional

$$J(v) = \frac{1}{\Delta t}\int_\Omega \frac{1}{2}|v|^2 + \frac{\eta}{2}\int_\Omega |Dv|^2 + \psi(v) - G(v). \tag{3.1.29}$$

This functional is the sum of the square of the H^{1s} norm and a l.s.c. function. Thus by Proposition 2.1.4(a), there is a unique minimum in $H_0^{1s}(\Omega)$. □

The Moreau envelope F_ε of F enables to build uniformly bounded approximations to our variational problem (3.1.28).

Lemma 3.1.8. *Assume that F satisfies Hypothesis 1, 2. Then for any $\varepsilon > 0$ there is a unique solution $u_\varepsilon \in H_0^{1s}(\Omega)$ to*

$$\forall v \in H_0^{1s}(\Omega) \quad \frac{1}{\Delta t}\int_\Omega u_\varepsilon \cdot (v - u_\varepsilon) + \int_\Omega F_\varepsilon(Dv) - \int_\Omega F_\varepsilon(Du_\varepsilon) + \eta \int_\Omega Du_\varepsilon : (Dv - Du_\varepsilon) \geq G(v - u_\varepsilon), \tag{3.1.30}$$

where G is a bounded linear form on $H_0^{1s}(\Omega)$, and η, Δt are positive constants. Moreover u_ε, Du_ε are bounded in $L^2(\Omega)$ uniformly in $\varepsilon \leq \varepsilon_0$.

Proof. According to Proposition 2.1.4 and Lemma 3.1.2(a), F_ε satisfies the same hypotheses 1, 2 as F. Thus applying Proposition 3.1.7 to F_ε we get the existence and uniqueness of u_ε. Then inserting $v = 0$ in (3.1.30) we get

$$\frac{1}{\Delta t}\int_\Omega |u_\varepsilon|^2 + \eta \int_\Omega |Du_\varepsilon|^2 + \int_\Omega F_\varepsilon(Du_\varepsilon) \leq G(u_\varepsilon) + \int_\Omega F_\varepsilon(0). \tag{3.1.31}$$

We have $F_\varepsilon(0) \leq F(0)$ which is finite according to Lemma 3.1.2. Following (3.1.3), F_ε is lower bounded by an affine function independent of $\varepsilon \leq \varepsilon_0$. The functional G is also bounded by a constant times the H^{1s} norm, thus from (3.1.31) we deduce bounds on u_ε and Du_ε in $L^2(\Omega)$ uniformly in $\varepsilon \leq \varepsilon_0$. □

Theorem 3.1.9 (Euler-Lagrange equations)**.** *Assume that F satisfies Hypothesis 1, 2, and let $\eta, \Delta t > 0$. Then finding $u \in H_0^{1s}(\Omega)$ solution to (3.1.28) where $G(v) = \langle f^n + \frac{u^n}{\Delta t}, v\rangle + \eta\langle Du^n, Dv\rangle$ for some $f^n \in L^2(\Omega)$, $u^n \in H^{1s}(\Omega)$, is equivalent to finding $(u, \sigma) \in H_0^{1s}(\Omega) \times L^2(\Omega)$ such that*

$$\begin{cases} \dfrac{u - u^n}{\Delta t} - \operatorname{div}\sigma - \eta \operatorname{div} Du = -\eta \operatorname{div} Du^n + f^n & \text{in } \Omega, \\ \sigma \in \partial F(Du) \quad \text{a.e. in } \Omega. \end{cases} \tag{3.1.32}$$

Moreover one has $F^*(\sigma) \in L^1(\Omega)$.

Proof. Assume that (3.1.32) holds, the first equation being understood in the sense of distributions. Then by density we can apply it to test functions in $H_0^{1s}(\Omega)$. Taking $v - u$ as test function, where $v \in H_0^{1s}(\Omega)$, we obtain denoting by $\langle ., . \rangle$ the $L^2(\Omega)$ scalar product,

$$\frac{1}{\Delta t}\langle u - u^n, v - u\rangle + \langle \sigma, Dv - Du\rangle + \eta\langle Du, Dv - Du\rangle = \eta\langle Du^n, Dv - Du\rangle + \langle f^n, v - u\rangle. \tag{3.1.33}$$

Since $\sigma \in \partial F(Du)$ a.e. we have $F(Dv) \geq F(Du) + \sigma : (Dv - Du)$ a.e. Thus

$$\int_\Omega F(Dv) \geq \int_\Omega F(Du) + \langle \sigma, Dv - Du\rangle. \tag{3.1.34}$$

Using this in (3.1.33) it follows that

$$\forall v \in H_0^{1s}(\Omega) \quad \frac{1}{\Delta t}\langle u, v - u\rangle + \int_\Omega F(Dv) - \int_\Omega F(Du) + \eta\langle Du, Dv - Du\rangle \geq G(v - u). \tag{3.1.35}$$

Hence (3.1.28) holds.

Reciprocally, assume that (3.1.28) holds. Following Proposition 2.1.4, we can consider the sequence of Moreau envelopes F_ε of F, which are convex and continuously differentiable and satisfy $F_\varepsilon \uparrow F$ pointwise as $\varepsilon \to 0$. We define u_ε as the unique solution to the problem in which we replace F by F_ε, i.e. u_ε solves (3.1.30). At fixed ε we take then $v = u_\varepsilon + sw$ in (3.1.30), where $w \in H_0^{1s}(\Omega)$ and s is a real number. Since F_ε is differentiable, we have

$$\frac{F_\varepsilon(Du_\varepsilon + sDw) - F_\varepsilon(Du_\varepsilon)}{s} \longrightarrow F'_\varepsilon(Du_\varepsilon) : Dw \quad \text{a.e.} \quad \text{as} \quad s \longrightarrow 0.$$

Moreover, since F'_ε is $1/\varepsilon$-Lipschitz, we also have for some $0 \le \theta_s(x) \le 1$

$$\begin{aligned}
\frac{F_\varepsilon(Du_\varepsilon + sDw) - F_\varepsilon(Du_\varepsilon)}{s} &= F'_\varepsilon(Du_\varepsilon + \theta_s sDw) : Dw, \\
\left|\frac{F_\varepsilon(Du_\varepsilon + sDw) - F_\varepsilon(Du_\varepsilon)}{s}\right| &\le \left(|F'_\varepsilon(0)| + \frac{1}{\varepsilon}|Du_\varepsilon + \theta_s sDw|\right)|Dw| \\
&\le \left(C + \frac{|Du_\varepsilon| + |s||Dw|}{\varepsilon}\right)|Dw|.
\end{aligned}$$

Therefore by Lebesgue's dominated convergence theorem,

$$\frac{F_\varepsilon(Du_\varepsilon + sDw) - F_\varepsilon(Du_\varepsilon)}{s} \longrightarrow F'_\varepsilon(Du_\varepsilon) : Dw \quad \text{in } L^1(\Omega) \quad \text{as} \quad s \to 0.$$

Taking $v = u_\varepsilon + sw$ in (3.1.30) where $s > 0$ and $w \in H_0^{1s}(\Omega)$ and dividing by s, one gets

$$\frac{1}{\Delta t}\langle u_\varepsilon, w\rangle + \int_\Omega \frac{F_\varepsilon(Du_\varepsilon + sDw) - F_\varepsilon(Du_\varepsilon)}{s} + \eta \int_\Omega Du_\varepsilon : Dw \ge G(w),$$

thus letting $s \to 0$

$$\frac{1}{\Delta t}\langle u_\varepsilon, w\rangle + \int_\Omega F'_\varepsilon(Du_\varepsilon) : Dw + \eta \int_\Omega Du_\varepsilon : Dw \ge G(w). \tag{3.1.36}$$

Similarly by taking $s < 0$ we get

$$\frac{1}{\Delta t}\langle u_\varepsilon, w\rangle + \int_\Omega F'_\varepsilon(Du_\varepsilon) : Dw + \eta \int_\Omega Du_\varepsilon : Dw \le G(w), \tag{3.1.37}$$

and it follows that

$$\frac{1}{\Delta t}\langle u_\varepsilon, w\rangle + \int_\Omega F'_\varepsilon(Du_\varepsilon) : Dw + \eta \int_\Omega Du_\varepsilon : Dw = G(w) \quad \forall w \in H_0^{1s}(\Omega). \tag{3.1.38}$$

Thus $(u_\varepsilon, \sigma_\varepsilon)$ is a solution to

$$\frac{u_\varepsilon - u^n}{\Delta t} - \operatorname{div} \sigma_\varepsilon - \eta \operatorname{div} Du_\varepsilon = -\eta \operatorname{div} Du^n + f^n, \tag{3.1.39}$$

$$\sigma_\varepsilon = F'_\varepsilon(Du_\varepsilon). \tag{3.1.40}$$

In order to get (3.1.32) it remains to pass to the limit as $\varepsilon \to 0$. According to Lemma 3.1.8, u_ε and Du_ε are bounded in $L^2(\Omega)$ uniformly in $\varepsilon \le \varepsilon_0$. Therefore with Lemma 3.1.2(a), σ_ε is bounded in L^2. Then according to Lemma 2.1.3(d) we have

$$Du_\varepsilon : \sigma_\varepsilon = F_\varepsilon(Du_\varepsilon) + (F_\varepsilon)^*(\sigma_\varepsilon). \tag{3.1.41}$$

According to Lemma 2.1.3(f), $(F_\varepsilon)^*$ is lower bounded by a constant independent of ε. Taking $w = u_\varepsilon$ in (3.1.38) we get that

$$\int_\Omega Du_\varepsilon : \sigma_\varepsilon \leq G(u_\varepsilon). \tag{3.1.42}$$

The right-hand side is bounded because u_ε and Du_ε are bounded in $L^2(\Omega)$, and with (3.1.41) we deduce that $\int_\Omega (F_\varepsilon)^*(\sigma_\varepsilon)$ is upper bounded. Since $(F_\varepsilon)^*$ is lower bounded by a constant we conclude that $(F_\varepsilon)^*(\sigma_\varepsilon)$ is bounded in $L^1(\Omega)$ independently of ε.

Hence after extraction of a subsequence, there exists $u \in H_0^{1s}(\Omega)$ and $\sigma \in L^2(\Omega)$ such that

$$u_\varepsilon \rightharpoonup u, \quad Du_\varepsilon \rightharpoonup Du, \quad \sigma_\varepsilon \rightharpoonup \sigma, \tag{3.1.43}$$

where $\rightharpoonup$ denotes the weak convergence in L^2, and additionally $u_\varepsilon \to u$ locally strongly. Letting $\varepsilon \to 0$ in (3.1.39), we obtain

$$\frac{u - u^n}{\Delta t} - \operatorname{div} \sigma - \eta \operatorname{div} Du = -\eta \operatorname{div} Du^n + f^n. \tag{3.1.44}$$

Now we would like to prove that $\sigma \in \partial F(Du)$ a.e. Since $\sigma_\varepsilon = F_\varepsilon'(Du_\varepsilon)$ a.e., we have

$$\forall W \quad F_\varepsilon(W) \geq F_\varepsilon(Du_\varepsilon) + \sigma_\varepsilon : (W - Du_\varepsilon) \quad \text{a.e.}$$

For $\varepsilon \leq \varepsilon_0$, we have $F_\varepsilon(Du_\varepsilon) \geq F_{\varepsilon_0}(Du_\varepsilon)$, thus

$$\forall W \quad F_\varepsilon(W) \geq F_{\varepsilon_0}(Du_\varepsilon) + \sigma_\varepsilon : (W - Du_\varepsilon) \quad \text{a.e.}$$

Multiplying by $\varphi \in C_c^\infty(\Omega)$, $\varphi \geq 0$, and taking the integral over Ω we obtain

$$\int_\Omega \varphi F_\varepsilon(W) \geq \int_\Omega \varphi F_{\varepsilon_0}(Du_\varepsilon) + \int_\Omega \varphi \sigma_\varepsilon : W - \int_\Omega \varphi \sigma_\varepsilon : Du_\varepsilon. \tag{3.1.45}$$

But multiplying (3.1.39) by φu_ε, we get

$$\int_\Omega \frac{u_\varepsilon}{\Delta t} \cdot u_\varepsilon \varphi + \int_\Omega \sigma_\varepsilon : D(\varphi u_\varepsilon) + \eta \int_\Omega Du_\varepsilon : D(\varphi u_\varepsilon) = G(\varphi u_\varepsilon),$$

$$\int_\Omega \varphi \frac{|u_\varepsilon|^2}{\Delta t} + \int_\Omega \varphi \sigma_\varepsilon : Du_\varepsilon + \int_\Omega \sigma_\varepsilon : (u_\varepsilon \otimes \nabla \varphi) + \eta \int_\Omega \varphi |Du_\varepsilon|^2 + \eta \int_\Omega Du_\varepsilon : (u_\varepsilon \otimes \nabla \varphi) = G(\varphi u_\varepsilon). \tag{3.1.46}$$

Similarly, taking φu as test function in (3.1.44), we get

$$\int_\Omega \varphi \frac{|u|^2}{\Delta t} + \int_\Omega \varphi \sigma : Du + \int_\Omega \sigma : (u \otimes \nabla \varphi) + \eta \int_\Omega \varphi |Du|^2 + \eta \int_\Omega Du : (u \otimes \nabla \varphi) = G(\varphi u). \tag{3.1.47}$$

Since $u_\varepsilon \to u$ locally strongly, we can pass to the limit in the terms involving $\nabla \varphi$. Thus taking the limit $\varepsilon \to 0$ in (3.1.46) and comparing to (3.1.47), we obtain

$$\int_\Omega \varphi \sigma_\varepsilon : Du_\varepsilon - \int_\Omega \varphi \sigma : Du + \eta \int_\Omega \varphi |Du_\varepsilon|^2 - \eta \int_\Omega \varphi |Du|^2 \longrightarrow 0.$$

Since $Du_\varepsilon \rightharpoonup Du$, we have $\int_\Omega \varphi |Du|^2 \leq \underline{\lim} \int_\Omega \varphi |Du_\varepsilon|^2$. Consequently we obtain

$$\overline{\lim} \int_\Omega \varphi \sigma_\varepsilon : Du_\varepsilon \leq \int_\Omega \varphi \sigma : Du.$$

Hence, taking the $\underline{\lim}$ as $\varepsilon \to 0$ in (3.1.45), we get

$$\int_\Omega \varphi F(W) \geq \int_\Omega \varphi F_{\varepsilon_0}(Du) + \int_\Omega \varphi\sigma : W - \int_\Omega \varphi\sigma : Du.$$

Then letting $\varepsilon_0 \to 0$ this yields

$$\int_\Omega \varphi F(W) \geq \int_\Omega \varphi F(Du) + \int_\Omega \varphi\sigma : W - \int_\Omega \varphi\sigma : Du.$$

This is true for all $\varphi \in C_c^\infty(\Omega)$ such that $\varphi \geq 0$, thus

$$F(W) \geq F(Du) + \sigma : W - \sigma : Du \quad \text{a.e.}$$

This is true for all W. But since F is finite everywhere and convex on a finite dimensional space, it is continuous, and consequently for a.e. $x \in \Omega$ the inequality holds for all W. We conclude that $\sigma \in \partial F(Du)$ a.e. in Ω. Therefore (u, σ) satisfies (3.1.32). We have already proved that this implies that u is the solution to (3.1.28), which is unique according to Proposition 3.1.7.
Finally the last assertion $F^*(\sigma) \in L^1(\Omega)$ follows from Lemma 2.1.3(d) that ensures that $\sigma : Du = F(Du) + F^*(\sigma)$ a.e. □

Remark: In the particular case of a Bingham fluid, i.e. $F(D) = |D|$, the proof of Euler-Lagrange equations was provided in [26, 58]. In this case, additionally to the regularization method that we have used here, the use of the Hahn-Banach theorem is possible. A stronger version of the Euler-Lagrange equations is also available for the Bingham problem, see [18].

3.2 Without viscosity

We consider now the case without viscosity (3.0.1), that corresponds to the unmodified problem (2.0.1). We shall need several growth assumptions on F.

Hypothesis 3. *F is lower bounded, which means that there is a constant E such that for all $D \in \mathbb{M}^s_{N\times N}(\mathbb{R})$*

$$E + F(D) \geq 0. \tag{3.2.1}$$

According to Lemma 2.1.3(f), this assumption is equivalent to $F^*(0) < \infty$.

Hypothesis 4. *F is finite everywhere and there is a constant $C > 0$ such that for all D*

$$F(D) \leq C(1 + |D|^2). \tag{3.2.2}$$

Hypothesis 5. *F is superlinear, which means that*

$$\frac{F(D)}{|D|} \longrightarrow \infty \quad \text{as} \quad |D| \longrightarrow \infty. \tag{3.2.3}$$

We consider again the Moreau envelope F_ε of F, as defined in Proposition 2.1.4. We have the following result.

Lemma 3.2.1. *Under Hypothesis 1, 3, 4, we have*
(a) F_ε is uniformly lower bounded,

$$E + F_\varepsilon(D) \geq 0, \tag{3.2.4}$$

(b) For all $\varepsilon > 0$ and all D,

$$F_\varepsilon(D) \leq C(1 + |D|^2), \tag{3.2.5}$$

(c) The estimates obtained in Lemma 3.1.3 are valid.
(d) If hypothesis 5 is satisfied, then it is also satisfied for F_ε, uniformly for $\varepsilon \leq \varepsilon_0$.

Proof. (a) By (3.2.1) one has $F(D) \geq -E$, and the definition (2.1.12) of F_ε gives immediately $F_\varepsilon(D) \geq -E$.
(b) One has $F_\varepsilon(D) \leq F(D)$, so that (3.2.2) gives immediately (3.2.5).
(c) Because of (3.2.1) and (3.2.2), F satisfies $|F(D)| \leq \max(C, E)(1+|D|^2)$. Therefore the assumptions of Lemma 3.1.3 are satisfied.
(d) Hypothesis 5 means that

$$\forall \lambda > 0, \quad \exists R > 0 \quad \inf_{|D| \geq R} \frac{F(D)}{|D|} \geq \lambda. \tag{3.2.6}$$

According to (2.1.12) one has

$$F_\varepsilon(D) = \inf_{\overline{D}} H_\varepsilon(\overline{D}, D), \tag{3.2.7}$$

with

$$H_\varepsilon(\overline{D}, D) = F(\overline{D}) + \frac{|\overline{D} - D|^2}{2\varepsilon}. \tag{3.2.8}$$

Let $\lambda > 0$ be given. Then according to (3.2.6) there exists $R > 0$ such that for all $|D| \geq R$, one has $F(D) \geq \lambda |D|$. Assuming that $\varepsilon \leq \varepsilon_0$ for some fixed $\varepsilon_0 > 0$, one can assume that $R > \lambda\varepsilon_0$. Then let us take D such that $|D| \geq R$. We have for all $\overline{D}$ satisfying $|\overline{D}| \geq R$,

$$H_\varepsilon(\overline{D}, D) \geq \lambda|\overline{D}| + \frac{|\overline{D} - D|^2}{2\varepsilon}. \tag{3.2.9}$$

The right-hand side is larger than its minimum over all $\overline{D}$, which is attained at $\overline{D} = (|D| - \lambda\varepsilon)_+ D/|D|$. Since $|D| \geq R > \lambda\varepsilon$, plugging this value in the right hand side of (3.2.9) yields

$$\forall |\overline{D}| \geq R, \quad H_\varepsilon(\overline{D}, D) \geq \lambda(|D| - \lambda\varepsilon) + \frac{\lambda^2\varepsilon}{2}. \tag{3.2.10}$$

For $|\overline{D}| \leq R$, one has

$$H_\varepsilon(\overline{D}, D) \geq -E + \frac{(|D| - R)^2}{2\varepsilon}. \tag{3.2.11}$$

We deduce that

$$\inf_{\overline{D}} H_\varepsilon(\overline{D}, D) \geq \min\Big(\lambda|D| - \lambda^2\varepsilon_0, -E + (|D| - R)^2/2\varepsilon_0\Big). \tag{3.2.12}$$

It follows with (3.2.7) that $F_\varepsilon(D)/|D| \geq \lambda/2$ as soon as $|D| \geq \overline{R}$, for some $\overline{R}$ large enough and depending only on $\lambda, \varepsilon_0, E, R$. This concludes that $F_\varepsilon(D)/|D| \to \infty$ as $|D| \to \infty$, uniformly in ε as long as $\varepsilon \leq \varepsilon_0$. □

We consider now an open bounded subset Ω of $\mathbb{R}^N$, and functions $u \in L^2(\Omega)$ that take values in $\mathbb{R}^N$. We use as previously the symmetric derivative $Du = (\nabla u + (\nabla u)^t)/2$ which is in general a distribution.

Definition 3.2.2. *Suppose that F satisfies Hypothesis 1, 3, 4, and define for Ω an open bounded subset of $\mathbb{R}^N$,*

$$K^F(\Omega) = \Big\{ u \in L^2(\Omega) \text{ such that } Du \in L^1(\Omega) \text{ and } F(Du) \in L^1(\Omega) \Big\}, \tag{3.2.13}$$

and $K_0^F(\Omega) \subset K^F(\Omega)$ by $u \in K_0^F(\Omega)$ if and only if

$$D\overline{u} = \overline{Du} \quad \text{in } \mathbb{R}^N, \tag{3.2.14}$$

where for a function w defined in Ω, $\overline{w}$ denotes the extension by 0 of w, i.e. the function defined over the whole space $\mathbb{R}^N$ as $\overline{w} = w$ in Ω, $\overline{w} = 0$ outside Ω.

Notice that since by Hypothesis 3 F is lower bounded by a constant $-E$, for any $u \in L^2(\Omega)$ such that $Du \in L^1(\Omega)$ one can consider $\int_\Omega F(Du) \in \overline{\mathbb{R}}$. Then this integral is finite if and only if $F(Du) \in L^1(\Omega)$. Note also that $C_c^\infty(\Omega) \subset K_0^F(\Omega) \subset K^F(\Omega)$.

Lemma 3.2.3. *When F satisfies Hypothesis 1, 3, 4, we have*
(a) $K^F(\Omega)$ and $K_0^F(\Omega)$ are convex and contain 0.
(b) Assuming additionally Hypothesis 5, if we have a sequence $u_n \in K^F(\Omega)$ such that $\|u_n\|_{L^2}$ and $\int_\Omega F(Du_n)$ are bounded, then $\|Du_n\|_{L^1}$ is bounded and there exist $u \in K^F(\Omega)$ and a subsequence $u_{n'}$ such that

$$u_{n'} \rightharpoonup u \text{ in } L^2(\Omega) \text{ weak}, \quad Du_{n'} \rightharpoonup Du \text{ in } L^1(\Omega) \text{ weak}. \tag{3.2.15}$$

Moreover we have then

$$\int_\Omega F(Du) \leq \underline{\lim} \int_\Omega F(Du_{n'}). \tag{3.2.16}$$

Additionally, if $u_n \in K_0^F(\Omega)$ then $u \in K_0^F(\Omega)$.

Proof. (a) If $u, v \in K^F(\Omega)$ and $0 \leq \theta \leq 1$, we have that $w = (1-\theta)u + \theta v \in L^2$, $Dw = (1-\theta)Du + \theta Dv \in L^1$, and by convexity of F

$$\int_\Omega F(Dw) \leq \int_\Omega \Big((1-\theta)F(Du) + \theta F(Dv) \Big) = (1-\theta) \int_\Omega F(Du) + \theta \int_\Omega F(Dv) < \infty. \tag{3.2.17}$$

Hence $w \in K^F(\Omega)$. This proves that $K^F(\Omega)$ is convex. Next if $u, v \in K_0^F(\Omega)$, by linearity of (3.2.14) one has also $w \in K_0^F(\Omega)$, proving the convexity of $K_0^F(\Omega)$. Finally $0 \in C_c^\infty(\Omega) \subset K_0^F(\Omega) \subset K^F(\Omega)$.
(b) Since F is superlinear (Hypothesis 5), there is some $R_0 > 0$ such that for any $R \geq R_0$, for all $|D| \geq R$ one has $|D| \leq K_R F(D)$, where $K_R > 0$ verifies $K_R \to 0$ as $R \to \infty$. Indeed one has just to take $\frac{1}{K_R} = \inf\limits_{|D| \geq R} \frac{F(D)}{|D|}$.
We deduce that Du_n is bounded in L^1. Indeed,

$$\begin{aligned}
\int\limits_\Omega |Du_n| &= \int\limits_{|Du_n| \geq R} |Du_n| + \int\limits_{|Du_n| < R} |Du_n| \\
&\leq K_R \int\limits_{|Du_n| \geq R} F(Du_n) + R|\Omega| \\
&= K_R \int_\Omega F(Du_n) - K_R \int\limits_{|Du_n| < R} F(Du_n) + R|\Omega| \\
&\leq K_R \int_\Omega F(Du_n) + K_R E|\{|Du_n| < R\}| + R|\Omega|,
\end{aligned} \tag{3.2.18}$$

and it is enough to take a single R to get the boundedness in L^1. Then the same estimate (3.2.18) gives that

$$\int\limits_{|Du_n|\geq R} |Du_n| \to 0, \qquad \text{as } R\to\infty, \quad \text{uniformly in } n, \tag{3.2.19}$$

which gives that Du_n is uniformly equi-integrable since Ω is bounded. Therefore we can extract a subsequence $u_{n'}$ such that $u_{n'} \rightharpoonup u$ in weak L^2, for some $u \in L^2$, and such that $Du_{n'} \rightharpoonup w$ weakly in L^1, where $w \in L^1$. This implies that $\langle Du_{n'}, \varphi\rangle \to \langle w, \varphi\rangle$ for all $\varphi \in C_c^\infty(\Omega)$. But since $u_{n'} \rightharpoonup u$ in weak L^2 we have $\langle Du_{n'}, \varphi\rangle = -\langle u_{n'}, \operatorname{div}\varphi\rangle \to -\langle u, \operatorname{div}\varphi\rangle$. It follows that $Du = w \in L^1$ and $Du_{n'} \rightharpoonup Du$ weakly in L^1.
One can check that the map which to $w(x) \in L^1(\Omega)$ associates $\int F(w)$ is convex, l.s.c. on L^1 strong. It follows that it is also l.s.c. on L^1-weak. Therefore, since $Du_{n'} \rightharpoonup Du$ in L^1-weak, we deduce that

$$\int\limits_\Omega F(Du) \leq \underline{\lim} \int\limits_\Omega F(Du_{n'}) < \infty.$$

This gives that $u \in K^F(\Omega)$ and concludes the proof in the case of $K^F(\Omega)$.
Finally, for the case $u_n \in K_0^F(\Omega)$, the convergences $u_{n'} \rightharpoonup u$ in weak $L^2(\Omega)$ and $Du_{n'} \rightharpoonup Du$ in weak $L^1(\Omega)$ imply that $\overline{u_{n'}} \rightharpoonup \overline{u}$ in weak $L^2(\mathbb{R}^N)$ and $\overline{Du_{n'}} \rightharpoonup \overline{Du}$ in weak $L^1(\mathbb{R}^N)$. Thus we can pass to the limit in (3.2.14) and obtain $u \in K_0^F(\Omega)$. □

Lemma 3.2.4. *Assume that F satisfies Hypothesis 1, 3, 4, and that Ω is strictly star-shaped. Then for $u \in K_0^F(\Omega)$ there exists a sequence $u_k \in C_c^\infty(\Omega)$ such that*

$$u_k \to u \text{ in } L^2(\Omega) \text{ strong}, \quad Du_k \to Du \text{ in } L^1(\Omega) \text{ strong}, \quad \int_\Omega F(Du_k) \to \int_\Omega F(Du). \tag{3.2.20}$$

Proof. The property for Ω to be star-shaped means that there exists a point $x_0 \in \Omega$ such that for any $x \in \Omega$, one has $[x_0, x] \subset \Omega$. It is strictly star-shaped if additionally for any $x \in \partial\Omega$, $[x_0, x) \subset \Omega$. Consider $u \in K_0^F(\Omega)$. Then (3.2.14) ensures that $\overline{u} \in L^2(\mathbb{R}^N)$, $D\overline{u} \in L^1(\mathbb{R}^N)$. Moreover

$$\int_{\mathbb{R}^N} \Big(F(D\overline{u}) - F(0)\Big) = \int_\Omega \Big(F(D\overline{u}) - F(0)\Big) = \int_\Omega F(Du) - |\Omega| F(0). \tag{3.2.21}$$

Consider then for $\lambda > 0$

$$v_\lambda(x) = \frac{1}{\lambda}\overline{u}(x_0 + \lambda(x - x_0)). \tag{3.2.22}$$

Then $v_\lambda \in L^2(\mathbb{R}^N)$, $Dv_\lambda \in L^1(\mathbb{R}^N)$ and as $\lambda \to 1$ one has $v_\lambda \to \overline{u}$ in $L^2(\mathbb{R}^N)$, $Dv_\lambda \to D\overline{u}$ in $L^1(\mathbb{R}^N)$, and

$$\begin{aligned} \int_{\mathbb{R}^N} \Big(F(Dv_\lambda) - F(0)\Big) &= \int_{\mathbb{R}^N} \Big(F\big(D\overline{u}(x_0 + \lambda(x - x_0))\big) - F(0)\Big) \\ &= \lambda^{-N} \int_{\mathbb{R}^N} \Big(F\big(D\overline{u}\big) - F(0)\Big). \end{aligned} \tag{3.2.23}$$

Consider then a smoothing sequence $\rho_\delta(x)$ on $\mathbb{R}^N$, and define $w_\delta = \rho_\delta * v_\lambda$, that indeed depends on δ and λ. Then we have that $w_\delta \in C_c^\infty(\mathbb{R}^N)$, and as $\delta \to 0$ at fixed λ, $w_\delta \to v_\lambda$ in $L^2(\mathbb{R}^N)$, $Dw_\delta \to Dv_\lambda$ in $L^1(\mathbb{R}^N)$. Moreover by Jensen's inequality

$$F(Dw_\delta(x)) = F\left(\int Dv_\lambda(x-y)\rho_\delta(y)dy\right) \leq \int F\big(Dv_\lambda(x-y)\big)\rho_\delta(y)dy = (\rho_\delta * F\big(Dv_\lambda\big))(x). \tag{3.2.24}$$

It follows that

$$\int_{\mathbb{R}^N}\Big(F(Dw_\delta)-F(0)\Big)\le\int_{\mathbb{R}^N}\Big(F(Dv_\lambda)-F(0)\Big). \tag{3.2.25}$$

Now, for a given integer n one can find $\lambda_n>1$ such that

$$\|v_{\lambda_n}-\overline{u}\|_{L^2(\mathbb{R}^N)}\le\frac{1}{n},\qquad \|Dv_{\lambda_n}-D\overline{u}\|_{L^1(\mathbb{R}^N)}\le\frac{1}{n}. \tag{3.2.26}$$

Define

$$\Omega_n=\Big\{x\in\mathbb{R}^N \text{ such that } x_0+\lambda_n(x-x_0)\in\Omega\Big\}\subset\Omega. \tag{3.2.27}$$

Since Ω is strictly star-shaped, one has $\partial\Omega_n\cap\partial\Omega=\emptyset$. Furthermore, v_{λ_n} has support in $\overline{\Omega_n}$, and $\operatorname{dist}(\overline{\Omega_n},\partial\Omega)>0$. Therefore for δ small enough, $w_\delta\in C_c^\infty(\Omega)$. Thus we can find such small $\delta=\delta_n$ so that

$$\|w_{\delta_n}-v_{\lambda_n}\|_{L^2(\mathbb{R}^N)}\le\frac{1}{n},\qquad \|Dw_{\delta_n}-Dv_{\lambda_n}\|_{L^1(\mathbb{R}^N)}\le\frac{1}{n}. \tag{3.2.28}$$

It follows that

$$\|w_{\delta_n}-\overline{u}\|_{L^2(\mathbb{R}^N)}\le\frac{2}{n},\qquad \|Dw_{\delta_n}-D\overline{u}\|_{L^1(\mathbb{R}^N)}\le\frac{2}{n}. \tag{3.2.29}$$

Moreover using (3.2.25), (3.2.23) and (3.2.21), one gets

$$\int_{\mathbb{R}^N}\Big(F(Dw_{\delta_n})-F(0)\Big)\le\lambda_n^{-N}\Big(\int_\Omega F(Du)-|\Omega|F(0)\Big). \tag{3.2.30}$$

Taking into account that both w_{δ_n} and $\overline{u}$ have support in Ω, (3.2.29) and the previous inequality yield that

$$w_{\delta_n}\to u \text{ in } L^2(\Omega),\quad Dw_{\delta_n}\to Du \text{ in } L^1(\Omega),\quad \overline{\lim}\int_\Omega F(Dw_{\delta_n})\le\int_\Omega F(Du). \tag{3.2.31}$$

But by lower semicontinuity one has

$$\int_\Omega F(Du)\le\underline{\lim}\int_\Omega F(Dw_{\delta_n}), \tag{3.2.32}$$

which concludes the proof. □

Definition 3.2.5. *Suppose that F satisfies Hypothesis 1, 3, 4, and define for Ω an open bounded subset of $\mathbb{R}^N$, $\psi:L^2(\Omega)\to\overline{\mathbb{R}}$ as*

$$\forall u\in L^2(\Omega),\quad \psi(u)=\begin{cases}\displaystyle\int_\Omega F(Du) & \text{if}\quad u\in K_0^F(\Omega),\\ \infty & \text{otherwise.}\end{cases} \tag{3.2.33}$$

Proposition 3.2.6. *Under Hypothesis 1, 3, 4, 5, we have that ψ is convex, proper and lower semi-continuous on $L^2(\Omega)$.*

Proof. (a) ψ is convex.
Assume that $u,v\in L^2(\Omega)$ and $\theta\in(0,1)$. If $u\notin K_0^F$ or $v\notin K_0^F$, then $\psi(u)=\infty$ or $\psi(v)=\infty$, and it follows obviously that $\psi((1-\theta)u+\theta v)\le\infty=(1-\theta)\psi(u)+\theta\psi(v)$. Next if $u\in K_0^F$ and $v\in K_0^F$

one has also $(1-\theta)u+\theta v \in K_0^F$ since K_0^F is convex, and $D((1-\theta)u+\theta v) = (1-\theta)Du+\theta Dv$. Then since F is convex

$$\psi((1-\theta)u+\theta v) = \int_\Omega F((1-\theta)Du+\theta Dv) \le \int_\Omega \Big((1-\theta)F(Du)+\theta F(Dv)\Big) = (1-\theta)\psi(u)+\theta\psi(v).$$

(b) ψ is proper.
Since F is finite everywhere one has $F(0)<\infty$, thus using that $0 \in K_0^F$ we deduce that $\psi(0) = |\Omega|F(0) < \infty$.
(c) ψ is l.s.c.
Suppose that $u_n \to u$ in $L^2(\Omega)$. We have to prove that $\psi(u) \le \underline{\lim}\,\psi(u_n)$. One can assume that $\underline{\lim}\,\psi(u_n) < \infty$, and extracting a subsequence if necessary, one can assume that $\psi(u_n)$ tends to some limit, that is necessarily finite since ψ is lower bounded. Then for n large enough $\psi(u_n)$ is finite, thus $u_n \in K_0^F(\Omega)$. Applying Lemma 3.2.3(b) we deduce that $u \in K_0^F(\Omega)$ and (3.2.16) yields that $\psi(u) \le \underline{\lim}\,\psi(u_n)$. □

Proposition 3.2.7. *Assume Hypothesis 1, 3, 4, 5 and consider $\Delta t > 0$, $u^n, f^n \in L^2(\Omega)$. Then there is one and only one solution $u \in L^2(\Omega)$ to the problem*

$$\forall v \in L^2(\Omega) \quad \frac{1}{\Delta t}\langle u, v-u\rangle + \psi(v) \ge \psi(u) + G(v-u), \tag{3.2.34}$$

which moreover satisfies

$$u = \operatorname*{argmin}_{v\in L^2(\Omega)} J(v), \tag{3.2.35}$$

where $J : L^2(\Omega) \to \overline{\mathbb{R}}$ is defined by

$$J(v) = \frac{1}{\Delta t}\int_\Omega \frac{|v|^2}{2} + \psi(v) - G(v), \tag{3.2.36}$$

with $G(v) = \langle f^n + \dfrac{u^n}{\Delta t}, v\rangle$, and ψ is as in Definition 3.2.5.

Proof. By Proposition 3.2.6 ψ is convex, proper and l.s.c. The result is thus classical, as in Proposition 2.1.4(a). □

Lemma 3.2.8. *Assume Hypothesis 1, 3, 4, 5, and that Ω is strictly star-shaped. Consider $\sigma \in L^2(\Omega)$ satisfying $\operatorname{div}\sigma \in L^2(\Omega)$ and $F^*(\sigma) \in L^1(\Omega)$. Then we have*

$$\forall w \in K_0^F(\Omega) \qquad \int_\Omega \sigma : Dw \ge -\int_\Omega (\operatorname{div}\sigma)\cdot w. \tag{3.2.37}$$

The meaning of the integral $\int \sigma : Dw$ in (3.2.37) needs to be explained. Under the assumptions on F, both F and F^* are lower bounded by a constant. Thus integrals like $\int F(Dw)$ or $\int F^*(\sigma)$ are well defined, as finite or $+\infty$. The properties $F(Dw) \in L^1$ (which holds since $w \in K_0^F(\Omega)$) or $F^*(\sigma) \in L^1$ (which is assumed here) mean exactly that these respective integrals are finite. Then according to the inequality (2.1.4) one has

$$\sigma : Dw \le F(Dw) + F^*(\sigma) \quad \text{a.e.} \tag{3.2.38}$$

It follows that $\sigma : Dw$ is upper bounded by an L^1 function, and thus that $\int \sigma : Dw$ is well-defined, as finite or $-\infty$. The equation (3.2.37) says in particular that this integral must be finite, i.e. $\sigma : Dw \in L^1(\Omega)$.

Proof of Lemma 3.2.8. Let $w \in K_0^F(\Omega)$. Then according to Lemma 3.2.4, there exists a sequence $w_k \in C_c^\infty(\Omega)$ such that $w_k \to w$ in L^2, $Dw_k \to Dw$ in L^1, and $\int F(Dw_k) \to \int F(Dw)$. Extracting if necessary a subsequence we have $Dw_k \to Dw$ a.e. in Ω. Since $w_k \in C_c^\infty$ one has

$$\int_\Omega \sigma : Dw_k = -\int_\Omega (\operatorname{div} \sigma) \cdot w_k. \tag{3.2.39}$$

Then according to the Fenchel-Young inequality we have

$$g_k \equiv F(Dw_k) + F^*(\sigma) - \sigma : Dw_k \geq 0 \quad \text{a.e. in } \Omega. \tag{3.2.40}$$

By assumption $F^*(\sigma) \in L^1$ thus $g_k \in L^1$. Since F is continuous we have that $g_k \to g$ a.e, with

$$g = F(Dw) + F^*(\sigma) - \sigma : Dw \geq 0. \tag{3.2.41}$$

By Fatou's Lemma we have

$$\int_\Omega g \leq \underline{\lim} \int_\Omega g_k. \tag{3.2.42}$$

Using (3.2.39) we have

$$\int_\Omega g_k = \int_\Omega F(Dw_k) + \int_\Omega F^*(\sigma) + \int_\Omega (\operatorname{div} \sigma) \cdot w_k, \tag{3.2.43}$$

thus

$$\underline{\lim} \int_\Omega g_k = \int_\Omega F^*(\sigma) + \int_\Omega (\operatorname{div} \sigma) \cdot w + \int_\Omega F(Dw) \tag{3.2.44}$$

Using this in (3.2.42) and taking into account the value (3.2.41) of g we obtain that $g \in L^1$, $\sigma : Dw \in L^1$ and

$$-\int_\Omega \sigma : Dw \leq \int_\Omega (\operatorname{div} \sigma) \cdot w, \tag{3.2.45}$$

which yields (3.2.37). □

Remark: If F is even, i.e. $F(-D) = F(D)$, then in the previous lemma we can apply the same result to $-w$ instead of w, thus we obtain that there is indeed equality in (3.2.37). In the general case, we do not know if there could be a strict inequality in (3.2.37).

Theorem 3.2.9 (Euler-Lagrange equations)**.** *With the same assumptions as in Proposition 3.2.7 and if Ω is strictly star-shaped, the solution $u \in L^2(\Omega)$ to the problem (3.2.34) is characterized by the existence of $\sigma \in L^2(\Omega)$ such that*

$$u \in K_0^F(\Omega), \;\; F^*(\sigma) \in L^1(\Omega), \tag{3.2.46}$$

$$\frac{u - u^n}{\Delta t} - \operatorname{div} \sigma = f^n \quad \text{in } \Omega, \tag{3.2.47}$$

$$\sigma \in \partial F(Du) \;\; \text{a.e. in } \Omega, \tag{3.2.48}$$

$$\int_\Omega \frac{u - u^n}{\Delta t} \cdot u + \int_\Omega F(Du) + \int_\Omega F^*(\sigma) \leq \int_\Omega f^n \cdot u. \tag{3.2.49}$$

Proof. Assume first that $u \in L^2$ and $\sigma \in L^2$ satisfy (3.2.46), (3.2.47), (3.2.48), (3.2.49). Then (3.2.47) gives that $\operatorname{div}\sigma \in L^2$, and according to Lemma 3.2.8 one has

$$\forall v \in K_0^F(\Omega) \qquad \int_\Omega \frac{u-u^n}{\Delta t}\cdot v + \int_\Omega \sigma : Dv \geq \int_\Omega f^n \cdot v. \tag{3.2.50}$$

Then according to (3.2.48) we have

$$\sigma : Du = F(Du) + F^*(\sigma) \quad \text{a.e. in } \Omega. \tag{3.2.51}$$

Replacing in (3.2.49) we obtain

$$\int_\Omega \frac{u-u^n}{\Delta t}\cdot u + \int_\Omega \sigma : Du \leq \int_\Omega f^n \cdot u. \tag{3.2.52}$$

Note that since $u \in K_0^F(\Omega)$, with (3.2.50) this yields that there is indeed equality in (3.2.52). Making the difference with (3.2.50) we obtain

$$\frac{1}{\Delta t}\langle u-u^n, v-u\rangle + \langle \sigma, Dv - Du\rangle \geq \langle f^n, v-u\rangle. \tag{3.2.53}$$

Since $\sigma \in \partial F(Du)$ a.e. we have $F(Dv) \geq F(Du) + \sigma : (Dv - Du)$ a.e. Thus

$$\int_\Omega F(Dv) \geq \int_\Omega F(Du) + \langle \sigma, Dv - Du\rangle. \tag{3.2.54}$$

Using this in (3.2.53) it follows that

$$\frac{1}{\Delta t}\langle u, v-u\rangle + \int_\Omega F(Dv) - \int_\Omega F(Du) \geq G(v-u). \tag{3.2.55}$$

Hence (3.2.34) holds for all $v \in K_0^F(\Omega)$. But if $v \in L^2(\Omega)$ and $v \notin K_0^F(\Omega)$, then $\psi(v) = \infty$, and (3.2.34) holds also trivially. Finally (3.2.34) holds for all $v \in L^2(\Omega)$.

Knowing that there is existence and uniqueness of a solution to (3.2.34), to conclude the theorem it remains only to prove that there is a solution $(u, \sigma) \in L^2 \times L^2$ to (3.2.46), (3.2.47), (3.2.48), (3.2.49).

According to Lemma 3.1.3(b), Hypothesis 2 is satisfied. Therefore for $\eta > 0$ we can apply Proposition 3.1.7 and Theorem 3.1.9. It gives $u_\eta \in H_0^{1s}(\Omega)$, $\sigma_\eta \in L^2(\Omega)$ satisfying

$$\frac{u_\eta - u^n}{\Delta t} - \operatorname{div}\sigma_\eta - \eta \operatorname{div} Du_\eta = f^n \qquad \text{in } \Omega, \tag{3.2.56}$$

$$\sigma_\eta \in \partial F(Du_\eta) \quad \text{a.e. in } \Omega, \tag{3.2.57}$$

and $F^*(\sigma_\eta) \in L^1(\Omega)$. Equation (3.2.56) can also be formulated as

$$\forall w \in H_0^{1s}(\Omega), \quad \frac{1}{\Delta t}\int_\Omega u_\eta \cdot w + \int_\Omega \sigma_\eta : Dw + \eta \int_\Omega Du_\eta : Dw = \int_\Omega \Big(f^n + \frac{u^n}{\Delta t}\Big)\cdot w. \tag{3.2.58}$$

Taking $w = u_\eta$ we get

$$\frac{1}{\Delta t}\int_\Omega |u_\eta|^2 + \eta \int_\Omega |Du_\eta|^2 + \int_\Omega \sigma_\eta : Du_\eta = \int_\Omega \Big(f^n + \frac{u^n}{\Delta t}\Big)\cdot u_\eta. \tag{3.2.59}$$

According to (3.2.57) and Lemma 2.1.3(d) (case of equality) we have

$$\sigma_\eta : Du_\eta = F(Du_\eta) + F^*(\sigma_\eta) \quad \text{a.e.} \tag{3.2.60}$$

Thus we can rewrite (3.2.59) as

$$\frac{1}{\Delta t}\int_\Omega |u_\eta|^2 + \eta\int_\Omega |Du_\eta|^2 + \int_\Omega \Big(E + F(Du_\eta)\Big) + \int_\Omega \Big(F(0) + F^*(\sigma_\eta)\Big) = \int_\Omega \Big(f^n + \frac{u^n}{\Delta t}\Big)\cdot u_\eta + |\Omega|(E + F(0)). \tag{3.2.61}$$

Writing that

$$\int_\Omega \Big(f^n + \frac{u^n}{\Delta t}\Big)\cdot u_\eta \le \frac{1}{2\Delta t}\int_\Omega |u_\eta|^2 + \frac{\Delta t}{2}\int_\Omega \Big|f^n + \frac{u^n}{\Delta t}\Big|^2,$$

we obtain the estimate

$$\frac{1}{2\Delta t}\int_\Omega |u_\eta|^2 + \eta\int_\Omega |Du_\eta|^2 + \int_\Omega \Big(E + F(Du_\eta)\Big) + \int_\Omega \Big(F(0) + F^*(\sigma_\eta)\Big) \le \frac{\Delta t}{2}\int_\Omega \Big|f^n + \frac{u^n}{\Delta t}\Big|^2 + |\Omega|(E + F(0)). \tag{3.2.62}$$

Since $F \ge -E$ and $F^* \ge -F(0)$ all the terms on the left-hand side are nonnegative, and we therefore get bounds on these quantities uniformly with respect to η. In particular we have

$$\|u_\eta\|_{L^2}^2 \le \Delta t^2\Big\|f^n + \frac{u^n}{\Delta t}\Big\|_{L^2}^2 + 2\Delta t|\Omega|(E + F(0)), \qquad \int_\Omega F(Du_\eta) \le \frac{\Delta t}{2}\Big\|f^n + \frac{u^n}{\Delta t}\Big\|_{L^2}^2 + |\Omega|F(0). \tag{3.2.63}$$

We claim that $u_\eta \in K_0^F(\Omega)$. Indeed since $u_\eta \in H_0^{1s}(\Omega)$, there exists a sequence $u^k \in C_c^\infty(\Omega)$ such that $u^k \to u_\eta$ in $L^2(\Omega)$ and $Du^k \to Du_\eta$ in $L^2(\Omega)$. Then $u^k \in K_0^F(\Omega)$, and passing to the limit in (3.2.14) it yields $u_\eta \in K_0^F(\Omega)$.

Applying now Lemma 3.2.3(b) we get the existence of $u \in K_0^F(\Omega)$ and a subsequence of ηs tending to 0 such that

$$u_\eta \rightharpoonup u \text{ in weak } L^2, \qquad Du_\eta \rightharpoonup Du \text{ in weak } L^1, \qquad \int F(Du) \le \underline{\lim} \int F(Du_\eta). \tag{3.2.64}$$

The estimate (3.2.62) also gives a bound on $\int F^*(\sigma_\eta)$ independent of η. By Lemma 3.1.3(a) we get a bound on $\|\sigma_\eta\|_{L^2}$. Therefore, extracting if necessary a subsequence, we obtain some $\sigma \in L^2(\Omega)$ such that

$$\sigma_\eta \rightharpoonup \sigma \quad \text{in } L^2 \text{ weak}. \tag{3.2.65}$$

Moreover we have also

$$\int F^*(\sigma) \le \underline{\lim} \int F^*(\sigma_\eta) < \infty. \tag{3.2.66}$$

According to (3.2.62) we have $\|Du_\eta\|_{L^2} \le C/\sqrt{\eta}$, and it enables to pass to the limit in the sense of distributions in (3.2.56), giving

$$\frac{u - u^n}{\Delta t} - \operatorname{div}\sigma = f^n \qquad \text{in } \Omega. \tag{3.2.67}$$

From (3.2.61) we obtain by lower semicontinuity

$$\frac{1}{\Delta t}\int_\Omega |u|^2 + \int_\Omega F(Du) + \int_\Omega F^*(\sigma) \le \int_\Omega \Big(f^n + \frac{u^n}{\Delta t}\Big)\cdot u. \tag{3.2.68}$$

Thus we have proved (3.2.46), (3.2.47), (3.2.49), and it only remains to prove (3.2.48), i.e. $\sigma \in \partial F(Du)$ a.e. Taking into account (3.2.67) we apply Lemma 3.2.8 to $w = u$, giving

$$\frac{1}{\Delta t}\int_\Omega |u|^2 + \int_\Omega \sigma : Du \ge \int_\Omega \Big(f^n + \frac{u^n}{\Delta t}\Big)\cdot u. \tag{3.2.69}$$

Alltogether with (3.2.68) we obtain

$$\int_\Omega \Big(F(Du) + F^*(\sigma) - \sigma : Du \Big) \leq 0. \tag{3.2.70}$$

Applying Lemma 2.1.3(d) we conclude that $F(Du) + F^*(\sigma) - \sigma : Du = 0$ a.e. and thus that $\sigma \in \partial F(Du)$ a.e., concluding the proof.

□

Remark: The condition (3.2.48), i.e. $\sigma \in \partial F(Du)$ a.e. could be removed in the statement of the previous theorem, since as the end of the proof shows it is a consequence of the other conditions. Notice that the solution u is unique, but σ associated to u may be nonunique.

Remark: For applications to granular flows, it is important not to assume any special dependency of $F(D)$ with respect to D, such as dependency only on $|D|$. In particular, the above proofs would simplify somehow for an even nonlinearity F, i.e. verifying $F(-D) = F(D)$, but we cannot assume that, as the examples of Section 2.6 show. The assumptions we have made on F are only on its asymptotic growth. Only the assumption of superlinearity Hypothesis 5 is a bit restrictive, since it excludes a linear behavior at infinity. For a 1-homogeneous function such as described in Section 2.5 one should be able to get well-posedness results with the approach of [14]. In this case Du could be a measure instead of an L^1 function, which modifies significantly the proofs. However it would be desirable to consider a nonlinearity F that can be superlinear in certain directions, and asymptotically linear in other directions. At the time being we do not know how to treat this case.

The superlinearity assumption can indeed be related to the finiteness of F^*, as the following lemma shows.

Lemma 3.2.10. *Let F be a convex, proper, l.s.c. function on a Hilbert space. If F is superlinear i.e. $\frac{F(D)}{|D|} \to \infty$ as $|D| \to \infty$, then F^* is finite everywhere. In finite dimension the converse is true.*

Proof. Suppose that F is superlinear. One has

$$F^*(\sigma) = \sup_D (\sigma : D - F(D)). \tag{3.2.71}$$

For a given σ, since F is superlinear there exist $M > 0$ (depending on $|\sigma|$) such that

$$|D| \geq M \Longrightarrow \frac{F(D)}{|D|} \geq |\sigma|. \tag{3.2.72}$$

Thus

$$\forall |D| \geq M, \qquad \sigma : D - F(D) \leq \sigma : D - |D||\sigma| \leq 0. \tag{3.2.73}$$

But since F is lower bounded by an affine function, there is some β such that $F(D) \geq \beta$ for all D such that $|D| \leq M$. It follows that

$$\forall |D| \leq M, \qquad \sigma : D - F(D) \leq M|\sigma| - \beta. \tag{3.2.74}$$

Therefore with (3.2.73) we deduce that $F^*(\sigma) < \infty$.

Conversely suppose that F is not superlinear. Then

$$\exists C > 0, \forall M > 0, \quad \exists |D| > M \text{ such that } F(D) \leq C|D|. \tag{3.2.75}$$

We can then find a sequence D_i such that $|D_i| \to \infty$ and $F(D_i) \leq C|D_i|$ for all i. Then we have for all σ

$$F^*(\sigma) = \sup_D(\sigma : D - F(D)) \geq \sup_i(\sigma : D_i - C|D_i|). \tag{3.2.76}$$

We can find a subsequence such that $\frac{D_i}{|D_i|} \rightharpoonup V$, for some unit vector V (here we use that we are in finite dimension, otherwise V could be null). Then for σ such that $|\sigma| > C$ and $\frac{\sigma}{|\sigma|} = V$, one has $F^*(\sigma) = \infty$, which finishes the proof. $\square$

Chapter 4

Explicit primal-dual algorithm

In this chapter we introduce a numerical scheme based on an algorithm used in image processing [20]. We give convergence estimates on both the space continuous formulation and the approximation by finite elements (FEM). The scheme is intended to solve the viscoplastic problem with or without viscosity, the well-posedness of which has been established in the previous chapter. Two algorithms are introduced, one is the first-order primal dual algorithm as in [20], the other is the acceleration scheme with the theoretical higher order convergence rate $O(1/n^2)$ whereas the former has the rate $O(1/n)$ in terms of the primal-dual gap, where n is the number of iterations.
According to Chambolle et al. [20], our inviscid minimization problem

$$\inf_{u\in H}\int_\Omega \left(\alpha\frac{|u|^2}{2}+F(Du)+\eta\frac{|Du|^2}{2}-f\cdot u\right)\mathrm{dx}, \tag{4.0.1}$$

with $\eta\geq 0$, can be considered formally as a particular case of the problem

$$\inf_{u\in H}\big(h(Ku)+g(u)\big), \tag{4.0.2}$$

where g,h are convex, proper, l.s.c. and $K:H\longrightarrow L^2(\Omega)$ is a linear map. For us

$$Ku=Du,\quad h(Ku)=\int_\Omega F(Du),\quad g(u)=\int_\Omega\Big(\alpha\frac{|u|^2}{2}+\eta\frac{|Du|^2}{2}-f\cdot u\Big). \tag{4.0.3}$$

As before denote by $\langle .,.\rangle$ the L^2 duality. The principle of the algorithm is to rewrite (4.0.2) as a primal problem $(\mathcal{P})$ and a dual problem $(\mathcal{D})$ as follows

$$(\mathcal{P})=\inf_{u\in H}\Big(h(Ku)+g(u)\Big)=\inf_{u\in H}\sup_{\sigma\in L^2}\Big(\langle\sigma,Ku\rangle-h^*(\sigma)+g(u)\Big) \tag{4.0.4}$$

$$=\inf_{u\in H}\sup_{\sigma\in L^2}\Big(\langle u,K^*\sigma\rangle-h^*(\sigma)+g(u)\Big)$$

$$\geq\sup_{\sigma\in L^2}\inf_{u\in H}\Big(\langle u,K^*\sigma\rangle-h^*(\sigma)+g(u)\Big)=\sup_{\sigma\in L^2}\Big(-g^*(-K^*\sigma)-h^*(\sigma)\Big)=(\mathcal{D}). \tag{4.0.5}$$

Here K^* denotes the adjoint operator of K. In our case of $Ku=Du$, we have $K^*\sigma=-\operatorname{div}\sigma$. This formulation corresponds to (2.3.1), (2.3.2), and here the Lagrangian is

$$\mathcal{L}(\sigma,u)=\langle\sigma,Ku\rangle-h^*(\sigma)+g(u). \tag{4.0.6}$$

The solution (u,σ) is characterized by $\partial_u\mathcal{L}(\sigma,u)\ni 0$ and $\partial_\sigma\mathcal{L}(\sigma,u)\ni 0$, which can be written

$$\partial g(u)\ni -K^*\sigma,\qquad \partial h(Ku)\ni\sigma. \tag{4.0.7}$$

Then given u and σ, a way to measure the error from the exact solution is to compute the primal-dual gap

$$\begin{aligned}&h(Ku)+g(u)+g^*(-K^*\sigma)+h^*(\sigma)\\ =&\Big(g(u)+g^*(-K^*\sigma)-\langle u,-K^*\sigma\rangle\Big)+\Big(h(Ku)+h^*(\sigma)-\langle\sigma,Ku\rangle\Big)\geq 0,\end{aligned} \tag{4.0.8}$$

that vanishes only when (u,σ) is a solution. We can notice that when the first relation of (4.0.7) is satisfied (which corresponds to the momentum conservation (3.2.47)), the primal-dual gap reduces to (3.2.49).

The idea of the algorithm is to characterize a solution (u,σ) by

$$\begin{cases}Ku\in\partial h^*(\sigma),\\ -K^*\sigma\in\partial g(u),\end{cases}\iff\begin{cases}\sigma\in\partial h(Ku),\\ u\in\partial g^*(-K^*\sigma),\end{cases}\iff\begin{cases}\sigma=(\mathrm{Id}+r\partial h^*)^{-1}(\sigma+rKu),\\ u=(\mathrm{Id}+\tau\partial g)^{-1}(u-\tau K^*\sigma),\end{cases} \tag{4.0.9}$$

where $r>0$, $\tau>0$ are two parameters. It leads us to the iterative algorithm of [20],

$$\begin{cases}\sigma_{k+1}=(\mathrm{Id}+r\partial h^*)^{-1}(\sigma_k+rK\bar{u}_k),\\ u_{k+1}=(\mathrm{Id}+\tau\partial g)^{-1}(u_k-\tau K^*\sigma_{k+1}),\\ \bar{u}_{k+1}=u_{k+1}+\theta(u_{k+1}-u_k),\end{cases} \tag{4.0.10}$$

for $k\geq 0$, where $\theta\in[0,1]$. According to [20] the stability condition is

$$r\tau\|K\|^2\leq 1. \tag{4.0.11}$$

The iteration is completed with initial values u_0, σ_0, and we set $u_{-1}=u_0$ so that applying the $\bar{u}$ formula of (4.0.10) also to $k=-1$ gives $\bar{u}_0=u_0$.

The main algorithm is for the choice $\theta=1$, then $\bar{u}_k=2u_k-u_{k-1}$.

4.1 Continuous formulation

Taking into account the definitions (4.0.3), the algorithm (4.0.10) with $\theta=1$ can be written

$$\begin{cases}\sigma_{k+1}=\mathbb{P}_r\big(\sigma_k+r(2Du_k-Du_{k-1})\big),\\ \dfrac{u_{k+1}-u^n}{\Delta t}-\operatorname{div}\sigma_{k+1}+\dfrac{u_{k+1}-u_k}{\tau}-\operatorname{div}\eta Du_{k+1}=-\operatorname{div}\eta Du^n+f^n,\end{cases} \tag{4.1.1}$$

where $\mathbb{P}_r$ is defined by (2.4.1). One can see the formal consistency of (4.1.1) with

$$\begin{cases}\dfrac{\hat{u}-u^n}{\Delta t}-\operatorname{div}\hat{\sigma}-\operatorname{div}\eta D\hat{u}=-\operatorname{div}\eta Du^n+f^n,\\ \hat{\sigma}\in\partial F(D\hat{u}),\end{cases} \tag{4.1.2}$$

since according to Lemma 2.4.1(c), $\hat{\sigma}\in\partial F(D\hat{u})$ if and only if $\hat{\sigma}=\mathbb{P}_r(\hat{\sigma}+rD\hat{u})$.

When the viscosity is positive $\eta>0$, at each iteration (4.1.1) we have to solve an elliptic problem. On the contrary in the inviscid case $\eta=0$ each iteration is fully explicit. The term $(u_{k+1}-u_k)/\tau$

can be thought as a relaxation term that enforces the convergence. One can consider also a more regularizing algorithm, replacing it by $-\operatorname{div}(Du_{k+1} - Du_k)/\tau$.

We define $H = L^2(\Omega)$ if $\eta = 0$, and $H = H_0^{1s}(\Omega)$ if $\eta > 0$, with the norm $\|v\|^2_{H_0^{1s}} = \|v\|^2_{L^2} + \|Dv\|^2_{L^2}$. According to Theorems 3.1.9 and 3.2.9, under some hypotheses on F and Ω there is a solution $(\hat{u}, \hat{\sigma}) \in H \times L^2(\Omega)$ to (4.1.2), that satisfies $F(Du) \in L^1(\Omega)$ and $F^*(\sigma) \in L^1(\Omega)$.

In order to get error estimates we consider that $v \mapsto Dv$ is a bounded operator, i.e. there exists a constant $L \geq 0$ such that

$$\|Dv\|_{L^2} \leq L\|v\|_{L^2}. \tag{4.1.3}$$

Of course this is not true for the inviscid case since then $H = L^2$, but when using discrete approximations this becomes true with a constant L depending on the approximation space. In the case of taking the relaxation term as $-\operatorname{div}(Du_{k+1} - Du_k)/\tau$ the assumption (4.1.3) can be simply replaced by the definition $L \equiv 1$. In order to understand the mechanism of the estimates, we state a formal result with the assumption (4.1.3).

Proposition 4.1.1 (Formal)**.** *If $r\tau L^2 \leq 1$, then the sequence (u_k, σ_k) defined by (4.1.1) verifies $u_k \to \hat{u}$ in L^2 as $k \to \infty$, where $(\hat{u}, \hat{\sigma})$ is the solution to (4.1.2). Moreover, if $\eta > 0$ then $Du_k \to D\hat{u}$ in L^2 as $k \to \infty$. If $r\tau L^2 < 1$ then σ_k is bounded in L^2, and $\sigma_{k+1} - \sigma_k \to 0$ in L^2.*

Remark: The convergence of σ_k to $\hat{\sigma}$ is a difficult issue related to the non-uniqueness of $\hat{\sigma}$ (as shown in [33] in the case of a Bingham fluid). However $\operatorname{div}\hat{\sigma}$ is unique.

Proof. We use the shorthand notation $\|\cdot\|$ for the norm in $L^2(\Omega)$. In the proof we temporary ignore the regularity of u_{k+1}, σ_{k+1} and their integrability. By the definition (2.4.1) of $\mathbb{P}_r$ we have $\sigma_{k+1} = (\mathrm{Id} + r\partial F^*)^{-1}(\sigma_k + r(2Du_k - Du_{k-1}))$, which is equivalent to

$$\partial F^*(\sigma_{k+1}) \ni \frac{\sigma_k - \sigma_{k+1}}{r} + 2Du_k - Du_{k-1}.$$

It follows that

$$\forall \sigma \in L^2 \qquad \int F^*(\sigma) \geq \int F^*(\sigma_{k+1}) + \langle \frac{\sigma_k - \sigma_{k+1}}{r} + 2Du_k - Du_{k-1}, \sigma - \sigma_{k+1}\rangle.$$

Using the identity $2a \cdot b = |a|^2 + |b|^2 - |a - b|^2$, one gets

$$\int F^*(\sigma) \geq \int F^*(\sigma_{k+1}) + \frac{\|\sigma_k - \sigma_{k+1}\|^2}{2r} + \frac{\|\sigma - \sigma_{k+1}\|^2}{2r} - \frac{\|\sigma_k - \sigma\|^2}{2r} + \langle 2Du_k - Du_{k-1}, \sigma - \sigma_{k+1}\rangle. \tag{4.1.4}$$

Multiplying both sides of the momentum equation of (4.1.1) by $u - u_{k+1}$ and taking the integral over Ω, we get for any $u \in H$

$$\begin{aligned}\langle \frac{u_{k+1} - u^n}{\Delta t}, u - u_{k+1}\rangle + \langle \sigma_{k+1}, Du - Du_{k+1}\rangle + \langle \frac{u_{k+1} - u_k}{\tau}, u - u_{k+1}\rangle - \langle f^n, u - u_{k+1}\rangle \\ - \eta\langle Du^n - Du_{k+1}, Du - Du_{k+1}\rangle = 0.\end{aligned}$$

Applying again the quadratic identity above, we deduce that for any $\sigma \in L^2$

$$\begin{aligned}&\frac{\|u\|^2 - \|u_{k+1}\|^2 - \|u - u_{k+1}\|^2}{2\Delta t} - \langle f^n - \operatorname{div}\eta Du^n + \frac{u^n}{\Delta t}, u - u_{k+1}\rangle + \langle \sigma_{k+1}, Du\rangle \\ &- \langle \sigma, Du_{k+1}\rangle + \langle \sigma - \sigma_{k+1}, Du_{k+1}\rangle - \frac{\|u_{k+1} - u_k\|^2}{2\tau} - \frac{\|u - u_{k+1}\|^2}{2\tau} + \frac{\|u - u_k\|^2}{2\tau} \\ &+ \frac{\eta}{2}\|Du\|^2 - \frac{\eta}{2}\|Du_{k+1}\|^2 - \frac{\eta}{2}\|Du - Du_{k+1}\|^2 = 0.\end{aligned}$$

Taking the opposite we obtain

$$\begin{aligned}
&\frac{\|u_{k+1}\|^2}{2\Delta t} + \frac{\eta}{2}\|Du_{k+1}\|^2 - \langle f^n - \operatorname{div}\eta Du^n + \frac{u^n}{\Delta t}, u_{k+1}\rangle + \langle \sigma, Du_{k+1}\rangle \\
&+ \frac{\|u-u_{k+1}\|^2}{2\Delta t} + \frac{\eta}{2}\|Du - Du_{k+1}\|^2 + \frac{\|u_{k+1}-u_k\|^2}{2\tau} + \frac{\|u-u_{k+1}\|^2}{2\tau} \\
&= \frac{\|u\|^2}{2\Delta t} + \frac{\eta}{2}\|Du\|^2 - \langle f^n - \operatorname{div}\eta Du^n + \frac{u^n}{\Delta t}, u\rangle + \langle \sigma_{k+1}, Du\rangle + \frac{\|u-u_k\|^2}{2\tau} + \langle \sigma - \sigma_{k+1}, Du_{k+1}\rangle.
\end{aligned} \tag{4.1.5}$$

Hence, adding (4.1.5) to (4.1.4) one gets for all $(u,\sigma) \in H \times L^2$

$$\begin{aligned}
&\frac{\|\sigma_k - \sigma\|^2}{2r} + \frac{\|u-u_k\|^2}{2\tau} \\
&\geq \left(\frac{\|u_{k+1}\|^2}{2\Delta t} + \frac{\eta}{2}\|Du_{k+1}\|^2 - \int F^*(\sigma) - \langle f^n - \operatorname{div}\eta Du^n + \frac{u^n}{\Delta t}, u_{k+1}\rangle + \langle \sigma, Du_{k+1}\rangle\right) \\
&- \left(\frac{\|u\|^2}{2\Delta t} + \frac{\eta}{2}\|Du\|^2 - \int F^*(\sigma_{k+1}) - \langle f^n - \operatorname{div}\eta Du^n + \frac{u^n}{\Delta t}, u\rangle + \langle \sigma_{k+1}, Du\rangle\right) \\
&+ \frac{\|u_{k+1}-u_k\|^2}{2\tau} + \frac{\|u-u_{k+1}\|^2}{2\tau} + \frac{\|\sigma_k - \sigma_{k+1}\|^2}{2r} + \frac{\|\sigma - \sigma_{k+1}\|^2}{2r} \\
&+ \frac{\|u-u_{k+1}\|^2}{2\Delta t} + \frac{\eta}{2}\|Du - Du_{k+1}\|^2 + \langle \sigma - \sigma_{k+1}, 2Du_k - Du_{k-1} - Du_{k+1}\rangle.
\end{aligned} \tag{4.1.6}$$

Besides, $(\hat{u}, \hat{\sigma})$ being the exact solution to (4.1.2), we have

$$\forall \sigma \in L^2 \quad \int F^*(\sigma) \geq \int F^*(\hat{\sigma}) + \langle D\hat{u}, \sigma - \hat{\sigma}\rangle, \tag{4.1.7}$$

and applying the same estimate as in (4.1.5),

$$\begin{aligned}
\forall u \quad &\frac{\|\hat{u}\|^2}{2\Delta t} + \frac{\eta}{2}\|D\hat{u}\|^2 - \langle f^n - \operatorname{div}\eta Du^n + \frac{u^n}{\Delta t}, \hat{u}\rangle + \langle \hat{\sigma}, D\hat{u} - Du\rangle + \frac{\|u-\hat{u}\|^2}{2\Delta t} + \frac{\eta}{2}\|Du - D\hat{u}\|^2 \\
=&\frac{\|u\|^2}{2\Delta t} + \frac{\eta}{2}\|Du\|^2 - \langle f^n - \operatorname{div}\eta Du^n + \frac{u^n}{\Delta t}, u\rangle.
\end{aligned} \tag{4.1.8}$$

From (4.1.7), (4.1.8) one has for all $(u,\sigma) \in H \times L^2$

$$\begin{aligned}
&\left(\frac{\|u\|^2}{2\Delta t} + \frac{\eta}{2}\|Du\|^2 - \int F^*(\hat{\sigma}) - \langle f^n - \operatorname{div}\eta Du^n + \frac{u^n}{\Delta t}, u\rangle + \langle \hat{\sigma}, Du\rangle\right) \\
&- \left(\frac{\|\hat{u}\|^2}{2\Delta t} + \frac{\eta}{2}\|D\hat{u}\|^2 - \int F^*(\sigma) - \langle f^n - \operatorname{div}\eta Du^n + \frac{u^n}{\Delta t}, \hat{u}\rangle + \langle \sigma, D\hat{u}\rangle\right) \\
&\geq \frac{\|u-\hat{u}\|^2}{2\Delta t} + \frac{\eta}{2}\|Du - D\hat{u}\|^2.
\end{aligned} \tag{4.1.9}$$

Substituting $\sigma = \hat{\sigma}$, $u = \hat{u}$ in (4.1.6), we can then subtract to (4.1.9) with $\sigma = \sigma_{k+1}$ and $u = u_{k+1}$, to obtain

$$\begin{aligned}
&\frac{\|\sigma_k - \hat{\sigma}\|^2}{2r} + \frac{\|u_k - \hat{u}\|^2}{2\tau} + \langle \sigma_k - \hat{\sigma}, Du_k - Du_{k-1}\rangle \\
&\geq \frac{\|\sigma_{k+1} - \hat{\sigma}\|^2}{2r} + \frac{\|u_{k+1} - \hat{u}\|^2}{2\tau} + \langle \sigma_{k+1} - \hat{\sigma}, Du_{k+1} - Du_k\rangle - \langle \sigma_{k+1} - \sigma_k, Du_k - Du_{k-1}\rangle \\
&+ \frac{\|\sigma_{k+1} - \sigma_k\|^2}{2r} + \frac{\|u_{k+1} - u_k\|^2}{2\tau} + \frac{\|u_{k+1} - \hat{u}\|^2}{\Delta t} + \eta\|Du_{k+1} - D\hat{u}\|^2.
\end{aligned} \tag{4.1.10}$$

Now we have according to the Young inequality, for any $\lambda > 0$,

$$\left|\langle \sigma_{k+1} - \sigma_k, Du_k - Du_{k-1}\rangle\right| \leq \lambda \frac{\|Du_k - Du_{k-1}\|^2}{2} + \frac{\|\sigma_{k+1} - \sigma_k\|^2}{2\lambda}. \tag{4.1.11}$$

Therefore

$$\begin{aligned}
&\frac{\|\sigma_k - \hat{\sigma}\|^2}{2r} + \frac{\|u_k - \hat{u}\|^2}{2\tau} + \lambda \frac{\|Du_k - Du_{k-1}\|^2}{2} + \langle \sigma_k - \hat{\sigma}, Du_k - Du_{k-1}\rangle \\
\geq &\frac{\|\sigma_{k+1} - \hat{\sigma}\|^2}{2r} + \frac{\|u_{k+1} - \hat{u}\|^2}{2\tau} + \lambda \frac{\|Du_{k+1} - Du_k\|^2}{2} + \langle \sigma_{k+1} - \hat{\sigma}, Du_{k+1} - Du_k\rangle \\
&+ \left(\frac{1}{2r} - \frac{1}{2\lambda}\right) \|\sigma_{k+1} - \sigma_k\|^2 + \frac{\|u_{k+1} - u_k\|^2}{2\tau} - \lambda \frac{\|Du_{k+1} - Du_k\|^2}{2} + \frac{\|u_{k+1} - \hat{u}\|^2}{\Delta t} + \eta \|Du_{k+1} - D\hat{u}\|^2.
\end{aligned} \tag{4.1.12}$$

Supposing $\lambda \geq r$, we set

$$a_k = \frac{\|\sigma_k - \hat{\sigma}\|^2}{2r} + \frac{\|u_k - \hat{u}\|^2}{2\tau} + \lambda \frac{\|Du_k - Du_{k-1}\|^2}{2} + \langle \sigma_k - \hat{\sigma}, Du_k - Du_{k-1}\rangle \geq 0. \tag{4.1.13}$$

Since by (4.1.3) we have $\|Du_{k+1} - Du_k\| \leq L\|u_{k+1} - u_k\|$, (4.1.12) yields

$$\begin{aligned}
a_k \geq a_{k+1} &+ \left(\frac{1}{2r} - \frac{1}{2\lambda}\right) \|\sigma_{k+1} - \sigma_k\|^2 + \left(\frac{1}{2\tau L^2} - \frac{\lambda}{2}\right) \|Du_{k+1} - Du_k\|^2 \\
&+ \frac{\|u_{k+1} - \hat{u}\|^2}{\Delta t} + \eta \|Du_{k+1} - D\hat{u}\|^2.
\end{aligned} \tag{4.1.14}$$

With the condition $\tau\lambda L^2 \leq 1$, or in other words

$$r \leq \lambda \leq \frac{1}{\tau L^2}, \tag{4.1.15}$$

we deduce that the sequence $\{a_k\}$ is nonincreasing, and it follows that

$$\sum_{k=0}^{\infty} \|u_{k+1} - \hat{u}\|^2 < \infty, \quad \eta \sum_{k=0}^{\infty} \|Du_{k+1} - D\hat{u}\|^2 < \infty. \tag{4.1.16}$$

Consequently $u_k \to \hat{u}$, and $Du_k \to D\hat{u}$ if $\eta > 0$.

If $r\tau L^2 < 1$ then one can take $\lambda > r$ satisfying (4.1.15), and according to the Young inequality one has

$$a_k \geq \left(\frac{1}{2r} - \frac{1}{2\lambda}\right) \|\sigma_k - \hat{\sigma}\|^2. \tag{4.1.17}$$

Since a_k is nonincreasing it is bounded, and it follows that $\|\sigma_k - \hat{\sigma}\|$ is bounded, thus σ_k is bounded in L^2. We also have from (4.1.14) that

$$\left(\frac{1}{2r} - \frac{1}{2\lambda}\right) \sum_{k=0}^{\infty} \|\sigma_{k+1} - \sigma_k\|^2 < \infty, \tag{4.1.18}$$

which implies that $\|\sigma_{k+1} - \sigma_k\| \to 0$. □

Remark: In the case of taking the relaxation term as $-\operatorname{div}(Du_{k+1} - Du_k)/\tau$, all the terms in $1/\tau$ in (4.1.12) are modified, the norms $\|v\|$ of a quantity v are replaced by $\|Dv\|$, thus the assumption (4.1.3) is not necessary, we have simply to take $L = 1$.

Remark: The result says nothing about the ˘2018best˘2019 choice of the parameters r, τ. Meanwhile the convergence rate depends a lot on this choice, this is discussed in Chapter 5. Another heuristic efficient approach is proposed in [30]. A particular feature of this primal˘2013dual algorithm is that it can be accelerated when the function g is uniformly convex. In our case $g(u) = \int_\Omega \left(\alpha\frac{|u|^2}{2} + \eta\frac{|Du|^2}{2} - f\cdot u\right)$ is α-uniformly convex (recall that $\alpha = 1/\Delta t$). Thus according to [20] one can get convergence at rate $O(1/k^2)$ for the primal-dual gap, which means $O(1/k)$ for the velocity error.

Proposition 4.1.2 (Formal)**.** *Define the accelerated algorithm as*

- ***Initialization****: Choose u_0, σ_0, r_0, τ_0 such that $r_0\tau_0 L^2 \le 1$, and set $u_{-1} = u_0$.*
- ***Iteration*** *($k \ge 0$): Update r_k, τ_k, u_k, σ_k as*

$$\begin{cases} \sigma_{k+1} = \mathbb{P}_{r_k}\Big(\sigma_k + r_k(Du_k + \theta_k(Du_k - Du_{k-1}))\Big), \\ \dfrac{u_{k+1} - u^n}{\Delta t} - \operatorname{div}\sigma_{k+1} + \dfrac{u_{k+1} - u_k}{\tau_k} - \operatorname{div}\eta Du_{k+1} = -\operatorname{div}\eta Du^n + f^n, \\ \tau_{k+1} = \theta_{k+1}\tau_k, \quad r_{k+1} = \dfrac{r_k}{\theta_{k+1}}, \quad \text{with} \quad \theta_{k+1} = \frac{1}{\sqrt{1+2\frac{\tau_k}{\Delta t}}}. \end{cases} \tag{4.1.19}$$

Then $u_k \to \hat{u}$ in L^2 as $k \to \infty$, with $\|u_k - \hat{u}\|_{L^2} = O(1/k)$, where $(\hat{u}, \hat{\sigma})$ is the solution to (4.1.2). Moreover, if $r_0\tau_0 L^2 < 1$, then σ_k is bounded in L^2 and $\|\sigma_{k+1} - \sigma_k\|_{L^2} \to 0$.

Proof. Note that the value of θ_0 is not defined, but does not matter since it appears as a factor of $Du_0 - Du_{-1} = 0$. We proceed as in the previous proposition. The inequality (4.1.10) becomes

$$\begin{aligned} \frac{\|\sigma_k - \hat{\sigma}\|^2}{2r_k} + \frac{\|u_k - \hat{u}\|^2}{2\tau_k} \ge \frac{\|\sigma_{k+1} - \hat{\sigma}\|^2}{2r_k} + \frac{\|u_{k+1} - \hat{u}\|^2}{2\tau_k} + \frac{\|\sigma_{k+1} - \sigma_k\|^2}{2r_k} + \frac{\|u_{k+1} - u_k\|^2}{2\tau_k} + \frac{\|u_{k+1} - \hat{u}\|^2}{\Delta t} \\ + \eta\|Du_{k+1} - D\hat{u}\|^2 + \big\langle \sigma_{k+1} - \hat{\sigma}, Du_{k+1} - (Du_k + \theta_k(Du_k - Du_{k-1}))\big\rangle. \end{aligned} \tag{4.1.20}$$

The last term can be decomposed as

$$\begin{aligned} &\big\langle \sigma_{k+1} - \hat{\sigma}, Du_{k+1} - (Du_k + \theta_k(Du_k - Du_{k-1}))\big\rangle \\ =&\langle \sigma_{k+1} - \hat{\sigma}, Du_{k+1} - Du_k\rangle - \theta_k\langle \sigma_k - \hat{\sigma}, Du_k - Du_{k-1}\rangle - \theta_k\langle \sigma_{k+1} - \sigma_k, Du_k - Du_{k-1}\rangle. \end{aligned} \tag{4.1.21}$$

According to the Young inequality, for any $\lambda_k > 0$ we have

$$\theta_k\Big|\langle \sigma_{k+1} - \sigma_k, Du_k - Du_{k-1}\rangle\Big| \le \lambda_k\theta_k^2\frac{\|Du_k - Du_{k-1}\|^2}{2} + \frac{\|\sigma_{k+1} - \sigma_k\|^2}{2\lambda_k}. \tag{4.1.22}$$

Hence

$$\begin{aligned} &\frac{\|\sigma_{k+1} - \hat{\sigma}\|^2}{2r_k} + \frac{\|u_{k+1} - \hat{u}\|^2}{2\tau_k} + \langle \sigma_{k+1} - \hat{\sigma}, Du_{k+1} - Du_k\rangle \\ &+ \left(\frac{1}{r_k} - \frac{1}{\lambda_k}\right)\frac{\|\sigma_{k+1} - \sigma_k\|^2}{2} + \frac{\|u_{k+1} - u_k\|^2}{2\tau_k} + \frac{\|u_{k+1} - \hat{u}\|^2}{\Delta t} + \eta\|Du_{k+1} - D\hat{u}\|^2 \\ &\le \frac{\|\sigma_k - \hat{\sigma}\|^2}{2r_k} + \frac{\|u_k - \hat{u}\|^2}{2\tau_k} + \theta_k\langle \sigma_k - \hat{\sigma}, Du_k - Du_{k-1}\rangle + \lambda_k\theta_k^2\frac{\|Du_k - Du_{k-1}\|^2}{2}. \end{aligned} \tag{4.1.23}$$

Dividing by τ_k and taking $\lambda_k \geq r_k$, we get

$$\frac{\|\sigma_{k+1}-\hat{\sigma}\|^2}{2r_k\tau_k} + \left(\frac{1}{2\tau_k^2} + \frac{1}{\Delta t\tau_k}\right)\|u_{k+1}-\hat{u}\|^2 + \frac{1}{\tau_k}\langle \sigma_{k+1}-\hat{\sigma}, Du_{k+1}-Du_k\rangle + \frac{\eta}{\tau_k}\|Du_{k+1}-D\hat{u}\|^2$$
$$+\left(\frac{1}{r_k}-\frac{1}{\lambda_k}\right)\frac{\|\sigma_{k+1}-\sigma_k\|^2}{2\tau_k} + \left(\frac{1}{2\tau_k^2}\|u_{k+1}-u_k\|^2 - \frac{\lambda_k}{2\tau_k}\|Du_{k+1}-Du_k\|^2\right) + \frac{\lambda_k}{\tau_k}\frac{\|Du_{k+1}-Du_k\|^2}{2}$$
$$\leq \frac{\|\sigma_k-\hat{\sigma}\|^2}{2r_k\tau_k} + \frac{\|u_k-\hat{u}\|^2}{2\tau_k^2} + \frac{\theta_k}{\tau_k}\langle \sigma_k-\hat{\sigma}, Du_k-Du_{k-1}\rangle + \frac{\lambda_k\theta_k^2}{\tau_k}\frac{\|Du_k-Du_{k-1}\|^2}{2}. \tag{4.1.24}$$

We notice that

$$r_{k+1}\tau_{k+1} = r_k\tau_k, \qquad \frac{r_k}{\tau_k} = \frac{r_{k+1}}{\tau_{k+1}}\theta_{k+1}^2, \qquad \frac{1}{\tau_k} = \frac{\theta_{k+1}}{\tau_{k+1}}, \tag{4.1.25}$$
$$\frac{1}{2\tau_k^2} + \frac{1}{\Delta t\tau_k} = \frac{1}{2\tau_k^2}\left(1+2\frac{\tau_k}{\Delta t}\right) = \frac{1}{2\tau_k^2\theta_{k+1}^2} = \frac{1}{2\tau_{k+1}^2},$$

and we define $\{\lambda_k\}_{k\geq 0}$ by $r_0 \leq \lambda_0 \leq 1/(\tau_0 L^2)$ and the update formula

$$\lambda_{k+1} = \lambda_k\frac{\tau_{k+1}}{\tau_k\theta_{k+1}^2}. \tag{4.1.26}$$

It follows that $\lambda_{k+1}/r_{k+1} = \lambda_k/r_k$, and $\lambda_k \geq r_k$ for all k. By setting

$$a_k = \frac{\|\sigma_k-\hat{\sigma}\|^2}{2r_k\tau_k} + \frac{\|u_k-\hat{u}\|^2}{2\tau_k^2} + \frac{\lambda_k\theta_k^2}{\tau_k}\frac{\|Du_k-Du_{k-1}\|^2}{2} + \frac{\theta_k}{\tau_k}\langle \sigma_k-\hat{\sigma}, Du_k-Du_{k-1}\rangle \geq 0, \tag{4.1.27}$$

using that $\|Du_{k+1}-Du_k\| \leq L\|u_{k+1}-u_k\|$ we obtain

$$a_{k+1} + \frac{\eta}{\tau_k}\|Du_{k+1}-D\hat{u}\|^2 + \left(\frac{1}{r_k}-\frac{1}{\lambda_k}\right)\frac{\|\sigma_{k+1}-\sigma_k\|^2}{2\tau_k} + \left(\frac{1}{\tau_k^2L^2}-\frac{\lambda_k}{\tau_k}\right)\frac{\|Du_{k+1}-Du_k\|^2}{2} \leq a_k. \tag{4.1.28}$$

Since $\lambda_k\tau_kL^2 = (\lambda_0/r_0)r_k\tau_kL^2 = \lambda_0\tau_0L^2 \leq 1$, we have that $\{a_k\}$ is a nonincreasing nonnegative sequence. As a consequence,

$$a_0 \geq a_k \geq \frac{\|u_k-\hat{u}\|^2}{2\tau_k^2}. \tag{4.1.29}$$

Therefore $u_k \to \hat{u}$ in L^2 at rate $O(\tau_k)$. According to [20] one has $\tau_k \sim \frac{\Delta t}{k}$ for large k, thus $u_k \to \hat{u}$ at rate $O(1/k)$.

If $r_0\tau_0L^2 < 1$, then we can take λ_0 such that $r_0 < \lambda_0 < 1/(\tau_0L^2)$. Since λ_k/r_k and $\lambda_k\tau_k$ are constant we have then $r_k < \lambda_k < 1/(\tau_kL^2)$ and

$$a_k \geq \left(\frac{1}{r_k}-\frac{1}{\lambda_k}\right)\frac{\|\sigma_k-\hat{\sigma}\|^2}{2\tau_k} + \frac{\|u_k-\hat{u}\|^2}{2\tau_k^2}. \tag{4.1.30}$$

Since $(1/r_k - 1/\lambda_k)/\tau_k$ is a positive constant, it follows that $\|\sigma_k-\hat{\sigma}\|$ is bounded, so that σ_k is bounded in L^2. Moreover from (4.1.28) we have

$$\sum_{k=0}^{\infty}\left(\frac{1}{r_k}-\frac{1}{\lambda_k}\right)\frac{\|\sigma_{k+1}-\sigma_k\|^2}{2\tau_k} < \infty, \tag{4.1.31}$$

which proves that $\|\sigma_{k+1}-\sigma_k\| \to 0$. □

Remark: Since $\tau_k \to 0$ as $k \to \infty$, we have $r_k \to \infty$ and $\theta_k \to 1$.
Remark: It is not clear how to choose τ_0, r_0 (satisfying $r_0\tau_0 \leq 1/L^2$) and k in order to reach a given accuracy at the lowest cost, especially if L is large, which is the case when a space discretisation comes into play as in the next subsection.

4.2 Finite element approximation

We suppose now that Ω is a polyhedral domain, and that we have a mesh $\mathcal{T}_h$ as described in Section 1.5. We define

$$V_h := \{v_h \in C(\overline{\Omega}) \text{ such that } v_h\,|_K \text{ is affine for each cell } K \in \mathcal{T}_h,\ v_h\,|_{\partial\Omega} = 0\}, \tag{4.2.1}$$

$$\Lambda_h := \{\sigma_h \in L^\infty(\Omega) \text{ such that } \sigma_h\,|_K \text{ is constant for each } K \in \mathcal{T}_h\}. \tag{4.2.2}$$

Following Theorem 3.1.9, we consider the problem of finding $(\hat{u}_h, \hat{\sigma}_h) \in V_h \times \Lambda_h$ such that

$$\int_\Omega \frac{\hat{u}_h - u^n}{\Delta t} \cdot v_h + \int_\Omega \hat{\sigma}_h : Dv_h + \eta \int_\Omega D\hat{u}_h : Dv_h = \eta \int_\Omega Du^n : Dv_h + \int_\Omega f^n \cdot v_h \quad \forall v_h \in V_h, \tag{4.2.3}$$

$$\hat{\sigma}_h \in \partial F(D\hat{u}_h) \text{ a.e. in } \Omega. \tag{4.2.4}$$

As in Chapter 3 we have to consider the integrated version, which is: find $\hat{u}_h \in V_h$ such that

$$\forall v_h \in V_h \quad \frac{1}{\Delta t}\langle \hat{u}_h - u^n, v_h - \hat{u}_h\rangle + \int_\Omega F(Dv_h) - \int_\Omega F(D\hat{u}_h) + \eta\langle D\hat{u}_h - Du^n, Dv_h - D\hat{u}_h\rangle \geq \langle f^n, v_h - \hat{u}_h\rangle, \tag{4.2.5}$$

where $\langle .,. \rangle$ still denotes the L^2 duality.

Proposition 4.2.1. *We assume that $\eta \geq 0$, and that F is convex and finite everywhere. Then there exists one and only one solution $\hat{u}_h$ to the problem (4.2.5), and moreover it satisfies*

$$\hat{u}_h = \operatorname*{argmin}_{v_h \in V_h} J(v_h), \tag{4.2.6}$$

where J is defined by

$$J(v) = \frac{1}{\Delta t}\int_\Omega \frac{1}{2}|v|^2 + \frac{\eta}{2}\int_\Omega |Dv|^2 + \int_\Omega F(Dv) - G(v), \tag{4.2.7}$$

with $G(v) = \langle f^n + \frac{u^n}{\Delta t}, v\rangle + \eta\langle Du^n, Dv\rangle$.

Proof. We note that $V_h \subset H_0^{1s}(\Omega)$. Moreover for $v_h \in V_h$, $Dv_h \in L^\infty$. Since F is convex and finite everywhere on a finite-dimensional space, it is continuous and bounded on bounded sets. It follows that $v_h \mapsto \int_\Omega F(Dv_h)$ is a finite valued convex function on V_h. Since V_h has finite dimension, this functional is thus continuous. Since G is a linear form on V_h, the result follows from the classical Proposition 2.1.4(a). □

Proposition 4.2.2. *We assume that $\eta \geq 0$, and that in the viscous case ($\eta > 0$) F satisfies Hypothesis 1, 2, 3, whereas in the inviscid case ($\eta = 0$), F satisfies Hypothesis 1, 3, 4, 5. We denote $H = H_0^{1s}(\Omega)$ if $\eta > 0$, $H = L^2(\Omega)$ if $\eta = 0$. Let $\hat{u} \in H$ be the solution to the problem (3.1.28) if $\eta > 0$, respectively (3.2.34) if $\eta = 0$, and $\hat{u}_h$ be the solution to the problem (4.2.5). Then*

$$\frac{\|\hat{u}_h - \hat{u}\|_{L^2}^2}{\Delta t} + \eta\|D\hat{u}_h - D\hat{u}\|_{L^2}^2 \leq 2\widehat{R}_{\hat{u}}, \tag{4.2.8}$$

and

$$\Big|\int_\Omega F(D\hat{u}_h) - \int_\Omega F(D\hat{u})\Big| \leq 6\sqrt{M\widehat{R}_{\hat{u}}}, \tag{4.2.9}$$

where

$$\widehat{R}_{\hat{u}} = \inf_{v_h \in V_h} \left(\sqrt{M} \Big(\frac{\|v_h - \hat{u}\|_{L^2}^2}{\Delta t} + \eta \|Dv_h - D\hat{u}\|_{L^2}^2 \Big)^{1/2} + \max\Big(0, \int_\Omega F(Dv_h) - \int_\Omega F(D\hat{u})\Big) \right), \quad (4.2.10)$$

$$M = 2|\Omega|(E + F(0)) + \Delta t \|f^n + \frac{u^n}{\Delta t}\|_{L^2}^2 + \eta \|Du^n\|_{L^2}^2. \quad (4.2.11)$$

Proof. According to Propositions 3.1.7 and 3.2.7, $\hat{u}$ is the unique function in H such that

$$\forall v \in H \quad \frac{1}{\Delta t}\langle \hat{u}, v - \hat{u}\rangle + \eta \langle D\hat{u}, Dv - D\hat{u}\rangle + \psi(v) \geq \psi(\hat{u}) + G(v - \hat{u}), \quad (4.2.12)$$

where ψ is defined on H either by (3.1.22) or by (3.2.33). Thus $\psi(v)$ is either $\int_\Omega F(Dv)$ if $\eta > 0$, or if $\eta = 0$, $\int_\Omega F(Dv)$ for $v \in K_0^F(\Omega)$, $+\infty$ if $v \notin K_0^F(\Omega)$. We have $V_h \subset H_0^{1s}(\Omega) \subset K_0^F(\Omega)$, and $\psi(v)$ is equal to $\int_\Omega F(Dv)$ for $v \in V_h$. The problem (4.2.12) is also equivalent to minimizing J over H, with

$$\forall v \in H, \quad J(v) = \frac{1}{\Delta t} \int_\Omega \frac{1}{2}|v|^2 + \frac{\eta}{2} \int_\Omega |Dv|^2 + \psi(v) - G(v). \quad (4.2.13)$$

This value of J reduces to (4.2.7) when $v \in V_h$. The problem (4.2.5) with solution $\hat{u}_h \in V_h$ can be written also

$$\forall v_h \in V_h \quad \frac{1}{\Delta t}\langle \hat{u}_h, v_h - \hat{u}_h\rangle + \eta \langle D\hat{u}_h, Dv_h - D\hat{u}_h\rangle + \psi(v_h) \geq \psi(\hat{u}_h) + G(v_h - \hat{u}_h). \quad (4.2.14)$$

In order to prove the estimates, we follow [14] and we denote $f = f^n + \frac{u^n}{\Delta t}$. Setting $v_h = 0$ in (4.2.14), we get

$$\frac{\|\hat{u}_h\|^2}{\Delta t} + \eta \|D\hat{u}_h\|^2 + \psi(\hat{u}_h) \leq \psi(0) + \langle f, \hat{u}_h\rangle + \eta \langle Du^n, D\hat{u}_h\rangle. \quad (4.2.15)$$

Using the Young inequality we obtain

$$\frac{\|\hat{u}_h\|^2}{2\Delta t} + \frac{\eta}{2}\|D\hat{u}_h\|^2 + \psi(\hat{u}_h) \leq \psi(0) + \frac{\Delta t}{2}\|f\|^2 + \frac{\eta}{2}\|Du^n\|^2. \quad (4.2.16)$$

Since F satisfies Hypothesis 3 i.e. $F(D) \geq -E$ for all $D \in \mathbb{M}^s_{N\times N}(\mathbb{R})$, we deduce

$$\frac{\|\hat{u}_h\|^2}{2\Delta t} + \frac{\eta}{2}\|D\hat{u}_h\|^2 \leq |\Omega|(E + F(0)) + \frac{\Delta t}{2}\|f\|^2 + \frac{\eta}{2}\|Du^n\|^2. \quad (4.2.17)$$

From (4.2.14), one has for all $v_h \in V_h$

$$\frac{1}{\Delta t}\langle \hat{u}_h, \hat{u} - \hat{u}_h\rangle + \eta \langle D\hat{u}_h, D\hat{u} - D\hat{u}_h\rangle + \psi(\hat{u}) - \psi(\hat{u}_h) + R(v_h) \geq \langle f, \hat{u} - \hat{u}_h\rangle + \eta \langle Du^n, D\hat{u} - D\hat{u}_h\rangle, \quad (4.2.18)$$

with

$$R(v_h) = \frac{1}{\Delta t}\langle \hat{u}_h, v_h - \hat{u}\rangle + \eta \langle D\hat{u}_h, Dv_h - D\hat{u}\rangle + \psi(v_h) - \psi(\hat{u}) - \langle f, v_h - \hat{u}\rangle - \eta \langle Du^n, Dv_h - D\hat{u}\rangle. \quad (4.2.19)$$

Hence, taking the infimum gives

$$\frac{1}{\Delta t}\langle \hat{u}_h, \hat{u} - \hat{u}_h\rangle + \eta \langle D\hat{u}_h, D\hat{u} - D\hat{u}_h\rangle + \psi(\hat{u}) - \psi(\hat{u}_h) + \inf_{v_h \in V_h} R(v_h) \geq \langle f, \hat{u} - \hat{u}_h\rangle + \eta \langle Du^n, D\hat{u} - D\hat{u}_h\rangle. \quad (4.2.20)$$

Taking $v = \hat{u}_h$ in (4.2.12), we get

$$\frac{1}{\Delta t}\langle \hat{u}, \hat{u}_h - \hat{u}\rangle + \eta\langle D\hat{u}, D\hat{u}_h - D\hat{u}\rangle + \psi(\hat{u}_h) - \psi(\hat{u}) \geq \langle f, \hat{u}_h - \hat{u}\rangle + \eta\langle Du^n, D\hat{u}_h - D\hat{u}\rangle. \quad (4.2.21)$$

Adding the inequalities (4.2.20) and (4.2.21) yields

$$\frac{\|\hat{u} - \hat{u}_h\|^2}{\Delta t} + \eta\|D\hat{u} - D\hat{u}_h\|^2 \leq \inf_{v_h \in V_h} R(v_h). \quad (4.2.22)$$

We notice that M in (4.2.11) is twice the right-hand side of (4.2.17), thus

$$\frac{\|\hat{u}_h\|^2}{\Delta t} + \eta\|D\hat{u}_h\|^2 \leq M. \quad (4.2.23)$$

Therefore we can estimate (4.2.19) as

$$\begin{aligned} R(v_h) &\leq \frac{1}{\Delta t}\|\hat{u}_h\|\,\|v_h - \hat{u}\| + \eta\|D\hat{u}_h\|\,\|Dv_h - D\hat{u}\| + \big(\psi(v_h) - \psi(\hat{u})\big)_+ \\ &\qquad + \|f\|\,\|v_h - \hat{u}\| + \eta\|Du^n\|\,\|Dv_h - D\hat{u}\| \\ &\leq \Big(\frac{\|\hat{u}_h\|^2}{\Delta t} + \eta\|D\hat{u}_h\|^2\Big)^{1/2}\Big(\frac{\|v_h - \hat{u}\|^2}{\Delta t} + \eta\|Dv_h - D\hat{u}\|^2\Big)^{1/2} \\ &\quad + \Big(\Delta t\|f\|^2 + \eta\|Du^n\|^2\Big)^{1/2}\Big(\frac{\|v_h - \hat{u}\|^2}{\Delta t} + \eta\|Dv_h - D\hat{u}\|^2\Big)^{1/2} + \big(\psi(v_h) - \psi(\hat{u})\big)_+ \\ &\leq 2\sqrt{M}\Big(\frac{\|v_h - \hat{u}\|^2}{\Delta t} + \eta\|Dv_h - D\hat{u}\|^2\Big)^{1/2} + \big(\psi(v_h) - \psi(\hat{u})\big)_+. \end{aligned} \quad (4.2.24)$$

Taking the infimum over $v_h \in V_h$ and taking into account that $\hat{u} \in K_0^F(\Omega)$ so that $\psi(\hat{u}) = \int F(D\hat{u})$, it yields

$$\inf_{v_h \in V_h} R(v_h) \leq 2\widehat{R}_{\hat{u}}. \quad (4.2.25)$$

With (4.2.22) it gives (4.2.8).
Next, according to (4.2.20) one has

$$\begin{aligned} \psi(\hat{u}_h) - \psi(\hat{u}) &\leq \langle f, \hat{u}_h - \hat{u}\rangle + \eta\langle Du^n, D\hat{u}_h - D\hat{u}\rangle + \frac{1}{\Delta t}\langle \hat{u}_h, \hat{u} - \hat{u}_h\rangle + \eta\langle D\hat{u}_h, D\hat{u} - D\hat{u}_h\rangle + \inf_{v_h \in V_h} R(v_h) \\ &\leq \left(\Big(\Delta t\|f\|^2 + \eta\|Du^n\|^2\Big)^{1/2} + \Big(\frac{\|\hat{u}_h\|^2}{\Delta t} + \eta\|D\hat{u}_h\|^2\Big)^{1/2}\right) \\ &\qquad \times \Big(\frac{\|\hat{u}_h - \hat{u}\|^2}{\Delta t} + \eta\|D\hat{u}_h - D\hat{u}\|^2\Big)^{1/2} + \inf_{v_h \in V_h} R(v_h) \\ &\leq 2\sqrt{M}\Big(\frac{\|\hat{u}_h - \hat{u}\|^2}{\Delta t} + \eta\|D\hat{u}_h - D\hat{u}\|^2\Big)^{1/2} + \inf_{v_h \in V_h} R(v_h). \end{aligned} \quad (4.2.26)$$

Similarly using (4.2.21),

$$\begin{aligned} \psi(\hat{u}) - \psi(\hat{u}_h) &\leq \frac{1}{\Delta t}\langle \hat{u}, \hat{u}_h - \hat{u}\rangle + \eta\langle D\hat{u}, D\hat{u}_h - D\hat{u}\rangle - \langle f, \hat{u}_h - \hat{u}\rangle - \eta\langle Du^n, D\hat{u}_h - D\hat{u}\rangle \\ &\leq \left(\Big(\Delta t\|f\|^2 + \eta\|Du^n\|^2\Big)^{1/2} + \Big(\frac{\|\hat{u}\|^2}{\Delta t} + \eta\|D\hat{u}\|^2\Big)^{1/2}\right) \\ &\qquad \times \Big(\frac{\|\hat{u}_h - \hat{u}\|^2}{\Delta t} + \eta\|D\hat{u}_h - D\hat{u}\|^2\Big)^{1/2}. \end{aligned} \quad (4.2.27)$$

But similarly as (4.2.17), taking $v = 0$ in (4.2.12) yields that $\hat{u}$ satisfies the same bound (4.2.23) as $\hat{u}_h$. Thus together with (4.2.26) we obtain, noticing that by (4.2.22) the infimum is nonnegative,

$$\big|\psi(\hat{u}_h) - \psi(\hat{u})\big| \le 2\sqrt{M}\Big(\frac{\|\hat{u}_h - \hat{u}\|^2}{\Delta t} + \eta\|D\hat{u}_h - D\hat{u}\|^2\Big)^{1/2} + \inf_{v_h \in V_h} R(v_h). \tag{4.2.28}$$

Taking into account (4.2.22) we deduce that

$$\big|\psi(\hat{u}_h) - \psi(\hat{u})\big| \le 2\sqrt{M \inf_{\hat{v}_h \in V_h} R(\hat{v}_h)} + \inf_{v_h \in V_h} R(v_h). \tag{4.2.29}$$

But taking $v_h = 0$ in (4.2.19) we have

$$\begin{aligned} \inf_{v_h \in V_h} R(v_h) &\le R(0) = -\frac{1}{\Delta t}\langle \hat{u}_h, \hat{u}\rangle - \eta\langle D\hat{u}_h, D\hat{u}\rangle + \psi(0) - \psi(\hat{u}) + \langle f, \hat{u}\rangle + \eta\langle Du^n, D\hat{u}\rangle \\ &\le M + |\Omega|(E + F(0)) + M \\ &\le \frac{5}{2}M. \end{aligned} \tag{4.2.30}$$

Therefore (4.2.29) yields, with (4.2.25)

$$\big|\psi(\hat{u}_h) - \psi(\hat{u})\big| \le \Big(2\sqrt{M} + \sqrt{\frac{5}{2}M}\Big)\sqrt{\inf_{v_h \in V_h} R(v_h)} \le \sqrt{2}\Big(2 + \sqrt{\frac{5}{2}}\Big)\sqrt{M\widehat{R}_{\hat{u}}} \le 6\sqrt{M\widehat{R}_{\hat{u}}}. \tag{4.2.31}$$

□

Corollary 4.2.3. *With the same assumptions as in Proposition 4.2.2 and if Ω is strictly star-shaped, when $h \to 0$ one has $\hat{u}_h \to \hat{u}$ in $H_0^{1s}(\Omega)$ if $\eta > 0$, or in $L^2(\Omega)$ if $\eta = 0$. Moreover $D\hat{u}_h \rightharpoonup D\hat{u}$ in weak $L^1(\Omega)$ and $\int_\Omega F(D\hat{u}_h) \to \int_\Omega F(D\hat{u})$.*

Proof. Since $\hat{u} \in H$ and $\hat{u} \in K_0^F(\Omega)$ with $\psi(\hat{u}) < \infty$ if $\eta = 0$, for any $\varepsilon > 0$ there exists $w \in C_c^\infty(\Omega)$ such that

$$\|w - \hat{u}\|_H \le \varepsilon \text{ and } \Big|\int F(Dw) - \int F(D\hat{u})\Big| \le \varepsilon. \tag{4.2.32}$$

Indeed if $\eta > 0$, by definition of H_0^{1s} we can find a sequence $w_k \in C_c^\infty$ such that $\|w_k - \hat{u}\|_{H_0^{1s}} \to 0$. Then since F is subquadratic one has that $\int F(Dw_k) \to \int F(D\hat{u})$, and the result follows. In the case $\eta = 0$, the result follows from Lemma 3.2.4. Then, w being given in C_c^∞, for h small enough there is a function $v_h \in V_h$ such that

$$\|v_h - w\|_{H^1(\Omega)} \le \varepsilon \text{ and } \Big|\int F(Dv_h) - \int F(Dw)\Big| \le \varepsilon, \tag{4.2.33}$$

because again $v \mapsto \int F(Dv)$ is continuous on $H^{1s}(\Omega)$. From (4.2.32), (4.2.33) we get that for h small enough we have a function $v_h \in V_h$ such that

$$\|v_h - \hat{u}\|_H \le 2\varepsilon \text{ and } \Big|\int F(Dv_h) - \int F(D\hat{u})\Big| \le 2\varepsilon. \tag{4.2.34}$$

This proves that $\widehat{R}_{\hat{u}} \to 0$ as $h \to 0$. With the estimates (4.2.8), (4.2.9) we conclude the convergence of $\hat{u}_h$ to $\hat{u}$ in H and the convergence of $\int F(D\hat{u}_h)$. About the weak L^1 convergence of $D\hat{u}_h$, if $\eta > 0$ this is obvious since we have convergence in L^2. If $\eta = 0$ we apply Lemma 3.2.3 and get that after extraction of a subsequence, $D\hat{u}_h$ converges in weak L^1. Since the limit is necessarily $D\hat{u}$, it proves that the convergence holds without extracting any subsequence. □

Remark: In Corollary 4.2.3 the convergence holds as $h \to 0$ at Δt fixed. The analysis in the case of small Δt for a time-dependent problem is slighly different and is done in [14], at least in the case when F attains its minumum at 0, i.e. $E + F(0) = 0$.

Proposition 4.2.4 (Discrete Euler-Lagrange equations)**.** *We assume that $\eta \geq 0$, and that F is convex, finite everywhere and lower bounded. Then the solution $\hat{u}_h \in V_h$ to the problem (4.2.5) obtained in Proposition 4.2.1 is characterized by the existence of $\hat{\sigma}_h \in \Lambda_h$ such that the local equations (4.2.3), (4.2.4) hold.*

Proof. Assume first that $(\hat{u}_h, \hat{\sigma}_h) \in V_h \times \Lambda_h$ satisfy (4.2.3), (4.2.4). Consider $w_h \in V_h$. Then by (4.2.4) one has

$$F(Dw_h) \geq F(D\hat{u}_h) + \hat{\sigma}_h : (Dw_h - D\hat{u}_h) \quad \text{a.e. in } \Omega. \tag{4.2.35}$$

Integrating over Ω we obtain

$$\int_\Omega F(Dw_h) \geq \int_\Omega F(D\hat{u}_h) + \langle \hat{\sigma}_h, Dw_h - D\hat{u}_h \rangle. \tag{4.2.36}$$

Taking in (4.2.3) $v_h = w_h - \hat{u}_h$ and using the previous inequality we get (4.2.5) with the test function w_h.

Conversely, knowing that there is existence and uniqueness for (4.2.5), we only have to prove that there exists a solution $(\hat{u}_h, \hat{\sigma}_h) \in V_h \times \Lambda_h$ to (4.2.3), (4.2.4). Let us consider first the case when F is continuously differentiable. Then we have a solution $\hat{u}_h$ to (4.2.5). Consider a test function $w_h \in V_h$, and take $v_h = \hat{u}_h + tw_h$ for $t \neq 0$, in (4.2.5). We obtain

$$\begin{aligned} \frac{1}{\Delta t}\langle \hat{u}_h - u^n, tw_h \rangle + \eta \langle D\hat{u}_h - Du^n, tDw_h \rangle \\ + \int_\Omega F(D\hat{u}_h + tDw_h) - \int_\Omega F(D\hat{u}_h) \geq \langle f^n, tw_h \rangle. \end{aligned} \tag{4.2.37}$$

Dividing by $t > 0$ and letting $t \to 0$ we obtain using Lebesgue's theorem since $D\hat{u}_h$, Dw_h belong to L^∞,

$$\frac{1}{\Delta t}\langle \hat{u}_h - u^n, w_h \rangle + \eta \langle D\hat{u}_h - Du^n, Dw_h \rangle + \int_\Omega F'(D\hat{u}_h) : Dw_h \geq \langle f^n, w_h \rangle. \tag{4.2.38}$$

Using the same argument for $t < 0$ we obtain the converse inequality, we we deduce that (4.2.38) is indeed an equality. Since $\hat{u}_h \in V_h$ is continuous and piecewise affine, $D\hat{u}_h$ is piecewise constant, i.e. $D\hat{u}_h \in \Lambda_h$. Setting $\hat{\sigma}_h = F'(D\hat{u}_h) \in \Lambda_h$, we obtain (4.2.3), (4.2.4).

Now in the general case when F is not differentiable, for any $\varepsilon > 0$ we can consider the Moreau envelope F_ε of F of Proposition 2.1.4. Then F_ε is continuously differentiable, and converges monotonically to F as $\varepsilon \to 0$. We can apply the existence result to F_ε, thus there exist $(\hat{u}_h^\varepsilon, \hat{\sigma}_h^\varepsilon) \in V_h \times \Lambda_h$ satisfying

$$\int_\Omega \frac{\hat{u}_h^\varepsilon - u^n}{\Delta t} \cdot v_h + \int_\Omega \hat{\sigma}_h^\varepsilon : Dv_h + \eta \int_\Omega D\hat{u}_h^\varepsilon : Dv_h = \eta \int_\Omega Du^n : Dv_h + \int_\Omega f^n \cdot v_h \quad \forall v_h \in V_h, \tag{4.2.39}$$

$$\hat{\sigma}_h^\varepsilon \in \partial F_\varepsilon(D\hat{u}_h^\varepsilon) \text{ a.e. in } \Omega. \tag{4.2.40}$$

We have that $\hat{u}_h^\varepsilon$ also satisfies (4.2.5) with F replaced by F_ε. Since F is lower bounded by a constant $-E$ one has that F_ε is lower bounded by the same constant. Thus taking $v_h = 0$ as test function, we can do the same estimate (4.2.17) as in Proposition 4.2.2, and get

$$\frac{\|\hat{u}_h^\varepsilon\|^2}{2\Delta t} + \frac{\eta}{2}\|D\hat{u}_h^\varepsilon\|^2 \leq |\Omega|(E + F_\varepsilon(0)) + \frac{\Delta t}{2}\|f\|^2 + \frac{\eta}{2}\|Du^n\|^2. \tag{4.2.41}$$

We have $F_\varepsilon(0) \leq F(0)$, thus $\hat{u}_h^\varepsilon$ is bounded in H independently of ε. But since $\hat{u}_h^\varepsilon \in V_h$ and V_h is finite dimensional, we deduce that $\hat{u}_h^\varepsilon$ is bounded in V_h, and thus that $D\hat{u}_h^\varepsilon$ is bounded in L^∞. Since $\partial F(D)$ remains bounded when D lies in a bounded set, using Lemma 3.1.1 and Proposition 2.1.4(c)(e) we deduce that $F_\varepsilon'(D)$ remains bounded when D lies in a bounded set, independently of $\varepsilon \leq \varepsilon_0$. It follows that $\hat{\sigma}_h^\varepsilon$ is bounded in L^∞ independently of ε. Extracting a subsequence if necessary, (V_h and Λ_h are finite dimensional), we get $\hat{u}_h^\varepsilon \to \hat{u}_h \in V_h$, $\hat{\sigma}_h^\varepsilon \to \hat{\sigma}_h \in \Lambda_h$ as $\varepsilon \to 0$. The spaces V_h and Λ_h being finite dimensional, we have thus $D\hat{u}_h^\varepsilon \to D\hat{u}_h$ in L^∞, $\hat{\sigma}_h^\varepsilon \to \hat{\sigma}_h$ in L^∞. Thus we can pass to the limit in (4.2.39) and get (4.2.3). Then according to (4.2.40) one has for all D

$$F_\varepsilon(D) \geq F_\varepsilon(D\hat{u}_h^\varepsilon) + \hat{\sigma}_h^\varepsilon : (D - D\hat{u}_h^\varepsilon) \qquad \text{a.e. in } \Omega. \tag{4.2.42}$$

The 'almost everywhere' means indeed that it holds on all cells $K \in \mathcal{T}_h$, since the quantities are constant on each cell. For $\varepsilon \leq \varepsilon_0$ we have $F_\varepsilon(D\hat{u}_h^\varepsilon) \geq F_{\varepsilon_0}(D\hat{u}_h^\varepsilon)$. Thus letting $\varepsilon \to 0$ we get

$$F(D) \geq F_{\varepsilon_0}(D\hat{u}_h) + \hat{\sigma}_h : (D - D\hat{u}_h) \qquad \text{a.e. in } \Omega. \tag{4.2.43}$$

Finally we let $\varepsilon_0 \to 0$ and get that $\hat{\sigma}_h \in \partial F(D\hat{u}_h)$ a.e., i.e. (4.2.4). □

We can now give a complement to Corollary (4.2.3).

Proposition 4.2.5. *With the same assumptions as in Proposition 4.2.2 and if Ω is strictly star-shaped, when $h \to 0$ one has $\hat{u}_h \to \hat{u}$ in H. According to Proposition 4.2.4 there exists some $\hat{\sigma}_h \in \Lambda_h$ such that the local equations (4.2.3), (4.2.4) hold. Then after extraction of a subsequence one has $\hat{\sigma}_h \rightharpoonup \hat{\sigma}$ in weak $L^2(\Omega)$, where $(\hat{u}, \hat{\sigma})$ is a solution to (3.1.32) if $\eta > 0$, or to (3.2.46), (3.2.47), (3.2.48), (3.2.49) if $\eta = 0$. Moreover one has $\int_\Omega F^*(\hat{\sigma}_h) \to \int_\Omega F^*(\hat{\sigma})$.*

Proof. Since $\hat{\sigma}_h$ satisfies (4.2.4), we have

$$\hat{\sigma}_h : D\hat{u}_h = F(D\hat{u}_h) + F^*(\hat{\sigma}_h) \quad \text{a.e. in } \Omega. \tag{4.2.44}$$

The functions involved in this identity are indeed constant in each cell $K \in \mathcal{T}_h$. If follows that $F^*(\hat{\sigma}_h) \in L^\infty(\Omega)$. Then taking $v_h = \hat{u}_h$ in (4.2.3) and using (4.2.44) we obtain

$$\frac{\|\hat{u}_h\|^2}{\Delta t} + \eta\|D\hat{u}_h\|^2 + \int_\Omega F(D\hat{u}_h) + \int_\Omega F^*(\hat{\sigma}_h) = \eta\langle Du^n, D\hat{u}_h\rangle + \langle f^n + \frac{u^n}{\Delta t}, \hat{u}_h\rangle. \tag{4.2.45}$$

Since F and F^* are lower bounded, this gives bounds on $\|\hat{u}_h\|_H$, $\int F(D\hat{u}_h)$, $\int F^*(\hat{\sigma}_h)$ independent of h. By Lemma 3.1.3(a) we deduce that $\hat{\sigma}_h$ is bounded in $L^2(\Omega)$ independently of h. Thus after extraction of a subsequence, there is some $\hat{\sigma} \in L^2(\Omega)$ such that $\hat{\sigma}_h \rightharpoonup \hat{\sigma}$ in weak L^2. Since $\sigma \mapsto \int F^*(\sigma)$ is convex and l.s.c. on L^2, thus it is also l.s.c. on weak L^2. Therefore passing to the limit in (4.2.45) and using Corollary (4.2.3) we obtain

$$\frac{\|\hat{u}\|^2}{\Delta t} + \eta\|D\hat{u}\|^2 + \int_\Omega F(D\hat{u}) + \int_\Omega F^*(\hat{\sigma}) \leq \eta\langle Du^n, D\hat{u}\rangle + \langle f^n + \frac{u^n}{\Delta t}, \hat{u}\rangle. \tag{4.2.46}$$

Thus if $\eta = 0$ we get (3.2.49) and (3.2.46). Next, for $\varphi \in C_c^\infty(\Omega)$ and $\varepsilon > 0$, as in Corollary (4.2.3), for h small enough there exists $v_h \in V_h$ such that $\|v_h - \varphi\|_{H^1} \leq \varepsilon$. The formulation (4.2.3) ensures then that for h small enough

$$\left| \int_\Omega \frac{\hat{u}_h - u^n}{\Delta t} \cdot \varphi + \int_\Omega \hat{\sigma}_h : D\varphi + \eta \int_\Omega D\hat{u}_h : D\varphi - \eta \int_\Omega Du^n : D\varphi - \int_\Omega f^n \cdot \varphi \right| \leq C\varepsilon, \tag{4.2.47}$$

where C is a constant independent of h and ε. Letting $h \to 0$ we get

$$\left| \int_\Omega \frac{\hat{u} - u^n}{\Delta t} \cdot \varphi + \int_\Omega \hat{\sigma} : D\varphi + \eta \int_\Omega D\hat{u} : D\varphi - \eta \int_\Omega Du^n : D\varphi - \int_\Omega f^n \cdot \varphi \right| \leq C\varepsilon. \tag{4.2.48}$$

Since this holds for any $\varepsilon > 0$ we conclude that

$$\frac{\hat{u} - u^n}{\Delta t} - \operatorname{div} \hat{\sigma} - \eta \operatorname{div} D\hat{u} = -\eta \operatorname{div} Du^n + f^n, \tag{4.2.49}$$

in the sense of distributions in Ω. This proves the first equation of (3.1.32) in the case $\eta > 0$ or (3.2.47) in the case $\eta = 0$. Then we have to prove the second equation of (3.1.32) in the case $\eta > 0$, or (3.2.48) in the case $\eta = 0$, which is any case writes $\hat{\sigma} \in \partial F(D\hat{u})$ a.e. in Ω. We have proved in Theorem 3.2.9 that for $\eta = 0$ this condition was consequence of the others, thus it is not necessary to prove it. In the case $\eta > 0$ we can use the same argument. By density of C_c^∞ in H_0^{1s} we can take $\hat{u}$ as test function in (4.2.49), giving

$$\frac{\|\hat{u}\|^2}{\Delta t} + \eta \|D\hat{u}\|^2 + \langle \hat{\sigma}, D\hat{u} \rangle = \eta \langle Du^n, D\hat{u} \rangle + \langle f^n + \frac{u^n}{\Delta t}, \hat{u} \rangle. \tag{4.2.50}$$

Comparing to (4.2.46) we deduce that

$$\int_\Omega F(D\hat{u}) + \int_\Omega F^*(\hat{\sigma}) \leq \langle \hat{\sigma}, D\hat{u} \rangle, \tag{4.2.51}$$

which proves that $\hat{\sigma} \in \partial F(D\hat{u})$ a.e. in Ω. Finally we have $F(D\hat{u}) + F^*(\hat{\sigma}) = \hat{\sigma} : D\hat{u}$ a.e. in Ω, and we deduce that there is equality in (4.2.46) (in the case $\eta = 0$ this was already proved in Theorem 3.2.9). Therefore comparing to (4.2.45) we conclude that $\int_\Omega F^*(\hat{\sigma}_h) \to \int_\Omega F^*(\hat{\sigma})$. □

Next we prove the convergence of the algorithm (4.1.1), corresponding to $\theta = 1$, applied at the discrete level. This means that we define a sequence $(u_h^k, \sigma_h^k) \in V_h \times \Lambda_h$ for all $k \geq 0$ by the iteration formula: for $k \geq 0$

$$\begin{cases} \sigma_h^{k+1} = \mathbb{P}_r\big(\sigma_h^k + r(2Du_h^k - Du_h^{k-1})\big) \quad \text{a.e. in } \Omega, \\ \langle \dfrac{u_h^{k+1} - u^n}{\Delta t}, v_h \rangle + \langle \sigma_h^{k+1}, Dv_h \rangle + \langle \dfrac{u_h^{k+1} - u_h^k}{\tau}, v_h \rangle + \eta \langle Du_h^{k+1} - Du^n, Dv_h \rangle = \langle f^n, v_h \rangle, \quad \forall v_h \in V_h, \end{cases} \tag{4.2.52}$$

where $\mathbb{P}_r$ is defined by (2.4.1) and $\langle ., . \rangle$ still denotes the L^2 scalar product. Here $r > 0$ and $\tau > 0$ are parameters. We observe that the first equation in (4.2.52) deals with data that are constant in each cell $K \in \mathcal{T}_h$. Thus this formula is applied independently in each cell. By the Riesz theorem the second equation determines a unique solution $u_h^{k+1} \in V_h$.
Given $u_h^0 \in V_h$ and $\sigma_h^0 \in \Lambda_h$ we set $u_h^{-1} = u_h^0$. Then the above iteration formulas define $(u_h^k, \sigma_h^k) \in V_h \times \Lambda_h$ for all $k \geq 0$. One can see the formal consistency with (4.2.3), (4.2.4).

The operator $v_h \mapsto Dv_h$ being linear, it is bounded on V_h since V_h is finite dimensional,

$$\|Dv_h\|_{L^2} \leq L_h \|v_h\|_{L^2} \qquad \forall v_h \in V_h. \tag{4.2.53}$$

Theorem 4.2.6. *We assume that $\eta \geq 0$, and that F is convex, finite everywhere and lower bounded. We denote by $\hat{u}_h \in V_h$ the solution to (4.2.5) obtained in Proposition 4.2.1, and we assume that*

$$r\tau L_h^2 \leq 1. \tag{4.2.54}$$

Then the sequence defined by the algorithm (4.2.52) verifies $u_h^k \to \hat{u}_h$ in V_h as $k \to \infty$. Moreover if there is strict inequality in (4.2.54) then σ_h^k is bounded in Λ_h, and up to extraction of a subsequence one has $\sigma_h^k \to \hat{\sigma}_h$ in Λ_h, where $(\hat{u}_h, \hat{\sigma}_h) \in V_h \times \Lambda_h$ solves (4.2.3), (4.2.4), as obtained in Proposition 4.2.4.

Proof. We proceed as exposed in the formal proof of Proposition 4.1.1. Due to the definition (2.4.1) of $\mathbb{P}_r$, one has

$$\partial F^*(\sigma_h^{k+1}) \ni \frac{\sigma_h^k - \sigma_h^{k+1}}{r} + 2Du_h^k - Du_h^{k-1} \quad \text{a.e. in } \Omega. \tag{4.2.55}$$

We deduce that $F^*(\sigma_h^{k+1}) < \infty$ and thus since it is piecewise constant, $F^*(\sigma_h^{k+1}) \in L^\infty$ and

$$\forall \sigma_h \in \Lambda_h \qquad \int F^*(\sigma_h) \geq \int F^*(\sigma_h^{k+1}) + \langle \frac{\sigma_h^k - \sigma_h^{k+1}}{r} + 2Du_h^k - Du_h^{k-1}, \sigma_h - \sigma_h^{k+1} \rangle. \tag{4.2.56}$$

Using the quadratic identity, it follows that for any $\sigma_h \in \Lambda_h$

$$\int F^*(\sigma_h) \geq \int F^*(\sigma_h^{k+1}) + \frac{\|\sigma_h^k - \sigma_h^{k+1}\|^2}{2r} + \frac{\|\sigma_h - \sigma_h^{k+1}\|^2}{2r} - \frac{\|\sigma_h^k - \sigma_h\|^2}{2r} + \langle 2Du_h^k - Du_h^{k-1}, \sigma_h - \sigma_h^{k+1} \rangle. \tag{4.2.57}$$

Taking $v_h = u_h - u_h^{k+1}$ in the momentum equation of (4.2.52) we get for any $u_h \in V_h$

$$\begin{aligned} \langle \frac{u_h^{k+1} - u^n}{\Delta t}, u_h - u_h^{k+1} \rangle + \langle \sigma_h^{k+1}, Du_h - Du_h^{k+1} \rangle + \langle \frac{u_h^{k+1} - u_h^k}{\tau}, u_h - u_h^{k+1} \rangle \\ + \eta \langle Du_h^{k+1} - Du^n, Du_h - Du_h^{k+1} \rangle - \langle f^n, u_h - u_h^{k+1} \rangle = 0. \end{aligned} \tag{4.2.58}$$

Then following the proof of Proposition 4.1.1 we obtain (4.1.6). Besides, according to Proposition 4.2.4 one can complete the solution $\hat{u}_h$ to (4.2.5) by $\hat{\sigma}_h \in \Lambda_h$ so that $(\hat{u}_h, \hat{\sigma}_h)$ solves (4.2.3), (4.2.4). Then since $\hat{\sigma}_h \in \partial F(D\hat{u}_h)$ a.e., one has $F^*(\hat{\sigma}_h) \in L^\infty$ and

$$\forall \sigma_h \in \Lambda_h \qquad \int F^*(\sigma_h) \geq \int F^*(\hat{\sigma}_h) + \langle D\hat{u}_h, \sigma_h - \hat{\sigma}_h \rangle. \tag{4.2.59}$$

Then we follow the proof of Proposition 4.1.1. We take some λ such that

$$r \leq \lambda \leq \frac{1}{\tau L_h^2}, \tag{4.2.60}$$

which is possible according to (4.2.54). Defining a_k as (4.1.13), we obtain the inequality (4.1.14), and the estimates (4.1.16). We deduce that $u_h^k \to \hat{u}_h$ in L^2. In the case of strict inequality $r\tau L_h^2 < 1$, one can choose λ with strict inequalities in (4.2.60). Then (4.1.17), (4.1.18) hold, and it follows that σ_h^k is bounded in L^2 and $\|\sigma_h^{k+1} - \sigma_h^k\| \to 0$ as $k \to \infty$. Since Λ_h is finite dimensional, after extraction of a subsequence one has $\sigma_h^k \to \hat{\sigma}_h \in \Lambda_h$ (that may be different from the one previously considered). Passing to the limit in (4.2.52) and using that $u_h^k \to \hat{u}_h$ and $\sigma_h^{k+1} - \sigma_h^k \to 0$ we obtain (4.2.3), (4.2.4). □

One can also consider the accelerated algorithm (4.1.19), that can be written as defining $(u_h^k, \sigma_h^k) \in V_h \times \Lambda_h$ by

- **Initialization**: Choose $u_h^0 \in V_h$, $\sigma_h^0 \in \Lambda_h$, r_0, τ_0 such that $r_0\tau_0 L_h^2 \leq 1$, and set $u_h^{-1} = u_h^0$.
- **Iteration** ($k \geq 0$): Update r_k, τ_k, u_h^k, σ_h^k as

$$\begin{cases} \sigma_h^{k+1} = \mathbb{P}_{r_k}\Big(\sigma_h^k + r_k(Du_h^k + \theta_k(Du_h^k - Du_h^{k-1}))\Big) \quad \text{a.e. in } \Omega, \\ \\ \langle \dfrac{u_h^{k+1} - u^n}{\Delta t}, v_h\rangle + \langle \sigma_h^{k+1}, Dv_h\rangle + \langle \dfrac{u_h^{k+1} - u_h^k}{\tau_k}, v_h\rangle + \eta\langle Du_h^{k+1} - Du^n, Dv_h\rangle = \langle f^n, v_h\rangle, \quad \forall v_h \in V_h, \\ \\ \tau_{k+1} = \theta_{k+1}\tau_k, \quad r_{k+1} = \dfrac{r_k}{\theta_{k+1}}, \quad \text{with} \quad \theta_{k+1} = \frac{1}{\sqrt{1+2\frac{\tau_k}{\Delta t}}}. \end{cases} \tag{4.2.61}$$

Theorem 4.2.7. *We assume that $\eta \geq 0$, and that F is convex, finite everywhere and lower bounded. We denote by $\hat{u}_h \in V_h$ the solution to (4.2.5) obtained in Proposition 4.2.1, and we assume that*

$$r_0\tau_0 L_h^2 \leq 1, \tag{4.2.62}$$

where L_h is such that (4.2.53) holds. Then the sequence defined by the algorithm (4.2.61) verifies $u_h^k \to \hat{u}_h$ as $k \to \infty$ with $\|u_h^k - \hat{u}_h\|_{L^2} = O(1/k)$. Moreover if there is strict inequality in (4.2.62) then σ_h^k is bounded in Λ_h, and up to extraction of a subsequence one has $\sigma_h^k \to \hat{\sigma}_h$ in Λ_h, where $(\hat{u}_h, \hat{\sigma}_h) \in V_h \times \Lambda_h$ solves (4.2.3), (4.2.4).

Proof. It is identical to the one of the previous theorem, taking into account the arguments in the proof of Proposition 4.1.2. □

Remark: For the inviscid case $\eta = 0$, we would like to use a really explicit algorithm, thus not having to invert the mass matrix in order to get u_h^{k+1} from the second line of (4.2.52). With the notation of Section 1.5 we thus introduce the scalar product over $V_h \times V_h$

$$\langle u_h, v_h\rangle_h = \int_\Omega \bar{u}_h \cdot \bar{v}_h. \tag{4.2.63}$$

We then use this scalar product instead of the L^2 scalar product on $V_h \times V_h$. In particular, the second line of (4.2.52) is changed into

$$\langle \frac{u_h^{k+1} - \bar{u}^n}{\Delta t}, v_h\rangle_h + \langle \sigma_h^{k+1}, Dv_h\rangle + \langle \frac{u_h^{k+1} - u_h^k}{\tau}, v_h\rangle_h + \eta\langle Du_h^{k+1} - Du^n, Dv_h\rangle = \langle \bar{f}^n, v_h\rangle_h, \quad \forall v_h \in V_h, \tag{4.2.64}$$

where $\bar{u}^n$ and $\bar{f}^n$ are approximations in V_h of u^n and f^n respectively. Note that we do not modify the scalar product of matrices: the terms involving D are unmodified. Then with this formulation there is no linear system to invert at all, each iteration is the multiplication of the vector of unknowns by a (sparse) matrix. The definition of L_h has to be modified however, as

$$\|Dv_h\|_{L^2} \leq L_h\langle v_h, v_h\rangle_h^{1/2}, \quad \forall v_h \in V_h. \tag{4.2.65}$$

Remark: It is not straightforward to use a higher order approximation in space, for example by taking $V_h = \mathcal{P}_2$ for u and $\Lambda_h = \mathcal{P}_1$ for σ, since the projection step (first line of (4.2.52)) would not operate in the discrete space Λ_h. Thus one would have to consider a further projection to Λ_h as

$$\langle \sigma_h^{k+1}, \sigma_h\rangle = \langle \mathbb{P}_r\big(\sigma_h^k + r(2Du_h^k - Du_h^{k-1})\big), \sigma_h\rangle, \quad \forall \sigma_h \in \Lambda_h. \tag{4.2.66}$$

Then we cannot ensure the condition $\hat{\sigma}_h \in \partial F(D\hat{u}_h)$ a.e. for the limit $(\hat{u}_h, \hat{\sigma}_h)$ obtained as $k \to \infty$.

Chapter 5

Numerical experiments

5.1 1D Bingham model with Euler transport

In the numerical experiments, additionally to the viscoplastic rheology we include transport (inertial) terms. With such terms, incompressible viscoplastic models were considered years ago, and lots of explicit or reference solutions are available, as well as laboratory experiments. One of the obstacles in the numerical evaluation of compressible models with Euler transport terms is the lack of explicit solutions. In this first test we propose an analytical solution for the one-dimensional compressible Euler equations with Bingham rheology. This will be used as a particular 1d solution to our 2d model. In the 1d model the unknown $(\rho(t,x), u(t,x))$ has to satisfy

$$\begin{cases} \partial_t \rho + \partial_x(\rho u) = 0 & \text{in} \quad (0,T)\times(0,L), \\ \partial_t(\rho u) + \partial_x(\rho u^2 + p(\rho)) - \partial_x\left(\sigma_0 \dfrac{\partial_x u}{|\partial_x u|}\right) = f & \text{in} \quad (0,T)\times(0,L), \\ u(t,0) = u(t,L) = 0 & \text{for } t \in (0,T), \\ \rho(0,x) = \rho_{ini}(x), \quad u(0,x) = u_{ini}(x) & \text{for } x \in (0,L), \end{cases} \tag{5.1.1}$$

where $p(\rho) = \dfrac{1}{2}\rho^2$, $\Omega = (0,L)$ and $f(t,x)$ is a given force term. We take $L = 4$, $\sigma_0 = 1$, $T = 1$ and we build an exact solution such that u has the form

$$u(t,x) = \begin{cases} tx & \text{if } 0 \le x \le 1, \\ t & \text{if } 1 \le x \le 3, \\ t(4-x) & \text{if } 3 \le x \le 4. \end{cases} \tag{5.1.2}$$

Indeed ρ and f will be derived from this choice. In order to find the density we look for the flow $X(s,t,x)$ that satisfies

$$\frac{dX}{ds} = u(s,X), \quad X(t,t,x) = x. \tag{5.1.3}$$

Classically, a function $V(t,x)$ satisfies the linear transport equation

$$\partial_t V + u\partial_x V = 0 \tag{5.1.4}$$

if and only if $V(t,x) = V_0(X(s=0,t,x))$. In our case of u given by (5.1.2) a (not so simple) computation gives for $s,t \geq 0$ and $0 \leq x \leq 4$

$$X(s,t,x) = \begin{cases} \min\big(xe^{\frac{s^2-t^2}{2}},1\big) + \min\big((\frac{s^2-t^2}{2} - \ln\frac{1}{x})_+,2\big) + \big(1 - \frac{1}{x}e^{2+(t^2-s^2)/2}\big)_+ & \text{if } 0 \leq x \leq 1, \\ \min\big(e^{\frac{s^2-t^2}{2}+x-1},1\big) + \min\Big(\big(x-1+\frac{s^2-t^2}{2}\big)_+,2\Big) + \big(1 - e^{3-x+(t^2-s^2)/2}\big)_+ & \text{if } 1 \leq x \leq 3, \\ \min\big(\frac{1}{4-x}e^{\frac{s^2-t^2}{2}+2},1\big) + \min\Big(\big(\frac{s^2-t^2}{2}+2-\ln(4-x)\big)_+,2\Big) + \big(1-(4-x)e^{\frac{t^2-s^2}{2}}\big)_+ & \text{if } 3 \leq x \leq 4. \end{cases}$$

Then by taking $s=0$ and assuming $t \leq 2$ we get

$$X(s=0,t,x) = \begin{cases} xe^{-\frac{t^2}{2}} & \text{if } 0 \leq x \leq 1, \\ e^{x-1-\frac{t^2}{2}} & \text{if } 1 \leq x \leq 1+\frac{t^2}{2}, \\ x - \frac{t^2}{2} & \text{if } 1+\frac{t^2}{2} \leq x \leq 3, \\ 3 - \frac{t^2}{2} - \ln(4-x) & \text{if } 3 \leq x \leq 4 - e^{-\frac{t^2}{2}}, \\ 4-(4-x)e^{\frac{t^2}{2}} & \text{if } 4 - e^{-\frac{t^2}{2}} \leq x \leq 4. \end{cases} \tag{5.1.5}$$

Taking the spatial derivative of (5.1.4), we get $\partial_t\rho + \partial_x(\rho u) = 0$ with $\rho = \partial_x V$. Thus $\rho(t,x)$ is given by $\rho(t,x) = \rho_0(X(s=0,t,x))\partial_x X(s=0,t,x)$. Taking $\rho_0(x) = 1$ we obtain

$$\rho(t,x) = \partial_x X(s=0,t,x) = \begin{cases} e^{-\frac{t^2}{2}} & \text{if } 0 < x < 1, \\ e^{x-1-\frac{t^2}{2}} & \text{if } 1 < x < 1+\frac{t^2}{2}, \\ 1 & \text{if } 1+\frac{t^2}{2} < x < 3, \\ \frac{1}{4-x} & \text{if } 3 < x < 4 - e^{-\frac{t^2}{2}}, \\ e^{\frac{t^2}{2}} & \text{if } 4 - e^{-\frac{t^2}{2}} < x < 4. \end{cases} \tag{5.1.6}$$

Then we compute

$$\partial_x u = \begin{cases} t & \text{if } 0 < x < 1, \\ 0 & \text{if } 1 < x < 3, \\ -t & \text{if } 3 < x < 4, \end{cases} \tag{5.1.7}$$

and we take

$$\operatorname{sgn}\,\partial_x u = \begin{cases} 1 & \text{if } 0 < x < 1, \\ 2-x & \text{if } 1 < x < 3, \\ -1 & \text{if } 3 < x < 4, \end{cases} \tag{5.1.8}$$

which implies that

$$\partial_x(\operatorname{sgn}\,\partial_x u) = \begin{cases} 0 & \text{if } 0 < x < 1, \\ -1 & \text{if } 1 < x < 3, \\ 0 & \text{if } 3 < x < 4. \end{cases} \tag{5.1.9}$$

Therefore f is finally deduced as

$$\partial_t(\rho u) + \partial_x(\rho u^2 + p(\rho)) - \partial_x(\operatorname{sgn}\,\partial_x u) = f(t,x) = \begin{cases} xe^{-\frac{t^2}{2}}(1+t^2) & \text{if } 0 < x < 1, \\ e^{x-1-\frac{t^2}{2}} + e^{2x-2-t^2} + \sigma_0 & \text{if } 1 < x < 1+\frac{t^2}{2}, \\ 1+\sigma_0 & \text{if } 1+\frac{t^2}{2} < x < 3, \\ 1 - t^2 + \frac{1}{(4-x)^3} & \text{if } 3 < x < 4 - e^{-\frac{t^2}{2}}, \\ e^{\frac{t^2}{2}}(4-x)(1-t^2) & \text{if } 4 - e^{-\frac{t^2}{2}} < x < 4. \end{cases} \tag{5.1.10}$$

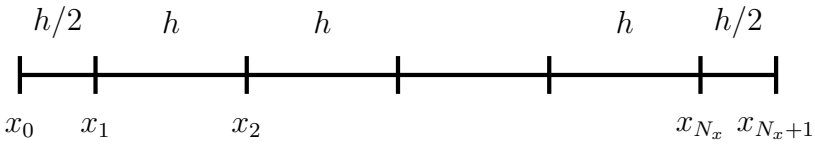


Figure 5.1: Locations x_i of the degrees of freedom and size of the cells

Space Discretisation. The interval $[0, L]$ is divided into N_x+1 finite element cells (x_0, x_1),...,(x_{N_x}, x_{N_x+1}) with

$$0 = x_0 < \frac{h}{2} = x_1 < \frac{3h}{2} = x_2 < \cdots < \left(N_x - \frac{1}{2}\right) h = x_{N_x} < L = x_{N_x+1}, \tag{5.1.11}$$

where $h = \dfrac{L}{N_x}$, see Figure 5.1. The two boundary cells have size half of the size of the internal cells. We denote $u_i^n \approx u(t_n, x_i)$, $\rho_i^n \approx \rho(t_n, x_i)$. In our scheme the unknown function $u(t, x)$ is reconstructed simultaneously from the values $u_h = (u_i)_{i\in\{0,\cdots,N_x+1\}}$ using a continuous piecewise affine reconstruction denoted by $\hat{u}_h$, and a piecewise constant reconstruction denoted by $\bar{u}_h$. Thus for all $u_h \in \mathbb{R}^{N_x+2}$ we have

$$\hat{u}_h \in C([0, L]), \hat{u}_h \text{ is affine for each interval } [x_i, x_{i+1}], \hat{u}_h(x_i) = u_i \quad \forall i \in \{0, \cdots, N_x + 1\}, \tag{5.1.12}$$

$$\bar{u}_h \in L^1_{loc}((0, L)), \bar{u}_h(x) = u_i \quad \forall x \in ((i-1)h, ih), i \in \{1, \cdots, N_x\}. \tag{5.1.13}$$

We use the same notation for ρ_h, $\bar{\rho}_h$.

Numerical scheme. The numerical scheme approximating (5.1.1) is given by

1- Initialisation of $u_h^0 \in \mathbb{R}^{N_x+2}$, $\rho_h^0 \in \mathbb{R}^{N_x+2}$:

$$\begin{aligned} u_i^0 &= u_{init}(x_i) \quad \forall i \in \{0, \cdots, N_x + 1\}, \\ \rho_i^0 &= \rho_{init}(x_i) \quad \forall i \in \{0, \cdots, N_x + 1\}. \end{aligned}$$

2- Finite volume step: Suppose that (u_h^n, ρ_h^n) are known. We define $u_h^{n+\frac{1}{2}}, \rho_h^{n+\frac{1}{2}}$ which approximate $(u^{n+\frac{1}{2}}, \rho^{n+\frac{1}{2}})$, the solution to 1.3.1:

$$U_i^{n+\frac{1}{2}} = U_i^n - \frac{\Delta t}{h}\left(F_{i+\frac{1}{2}} - F_{i-\frac{1}{2}}\right), \tag{5.1.14}$$

where $U_i^n := (\rho_i^n, \rho_i^n u_i^n)$ and similarly with $U_i^{n+\frac{1}{2}}$. We consider here only first-order explicit three points schemes where

$$F_{i+\frac{1}{2}} = F(U_i^n, U_{i+1}^n). \tag{5.1.15}$$

The function $F(U_l, U_r)$ is called the numerical flux. There are lots of different choices for F , some of being well-known and widely used, as the upwind, Lax-Friedrichs or Siliciu schemes. We will not discuss further this here, and just apply a known numerical flux such as the Suliciu one, see [11]. There is however an associated CFL condition (for Courant, Friedrichs, Levy [24]) on the timestep to prevent the blow up of the numerical values, under the form

$$\Delta t a \leq h, \tag{5.1.16}$$

where a is an approximation of the propagation speed.

We have to use compatible boundary conditions between the finite volume and finite element steps. We use the Dirichlet boundary condition in the finite element step and the "wall" condition

for the finite volume step. This means that, the basic unknowns being for $i = 1, \dots, N_x$, we set before the finite volume step

$$\rho_0^n = \rho_1^n, \qquad \rho_{N_x+1}^n = \rho_{N_x}^n, \qquad u_0^n = -u_1^n, \qquad u_{N_x+1}^n = -u_{N_x}^n. \tag{5.1.17}$$

3- Finite element step: for illustrative purpose we use the regularisation method. Suppose that $(u_h^{n+\frac{1}{2}}, \rho_h^{n+\frac{1}{2}})$ are known. We define u_h^{n+1}, ρ_h^{n+1} which approximate u^{n+1}, ρ^{n+1} the solution to (1.3.2), as $\rho_h^{n+1} = \rho_h^{n+1/2}$ and

$$\int_0^L \bar{\rho}_h^{n+1} \frac{\bar{u}_h^{n+1} - \bar{u}_h^{n+\frac{1}{2}}}{\Delta t} \bar{v}_h \, \mathrm{dx} + \sigma_0 \int_0^L \frac{\partial_x \hat{u}_h^{n+1}}{|\partial_x \hat{u}_h^{n+1}|} \partial_x \hat{v}_h \, \mathrm{dx} = \int_0^L \bar{f}_h \bar{v}_h \, \mathrm{dx} \quad \forall \hat{v}_h \in V_h. \tag{5.1.18}$$

Using the regularisation method, we replace this by the approximation, for a small positive ε,

$$\int_0^L \bar{\rho}_h^{n+1} \frac{\bar{u}_h^{n+1} - \bar{u}_h^{n+\frac{1}{2}}}{\Delta t} \bar{v}_h \, \mathrm{dx} + \sigma_0 \int_0^L \frac{\partial_x \hat{u}_h^{n+1}}{\sqrt{|\partial_x \hat{u}_h^{n+1}|^2 + \varepsilon^2}} \partial_x \hat{v}_h \, \mathrm{dx} = \int_0^L \bar{f}_h \bar{v}_h \, \mathrm{dx} \quad \forall \hat{v}_h \in V_h. \tag{5.1.19}$$

Note that ε is chosen "optimally" as

$$\varepsilon \sim 10^{-2} \frac{h^2 ||\partial_x u||^2 \rho}{\sigma_0} \frac{1}{T}, \tag{5.1.20}$$

see [43]. We use the fixed point method to get u_h^{n+1}. We initialize $u_h^{n+1,0} = u_h^{n+1/2}$, and when $u_h^{n+1,k}$ is known we compute $u_h^{n+1,k+1}$ by

$$\int_0^L \bar{\rho}_h^{n+1} \frac{\bar{u}_h^{n+1,k+1} - \bar{u}_h^{n+\frac{1}{2}}}{\Delta t} \bar{v}_h \, \mathrm{dx} + \sigma_0 \int_0^L \frac{\partial_x \hat{u}_h^{n+1,k+1}}{\sqrt{|\partial_x \hat{u}_h^{n+1,k}|^2 + \varepsilon^2}} \partial_x \hat{v}_h \, \mathrm{dx} = \int_0^L \bar{f}_h \bar{v}_h \, \mathrm{dx} \quad \forall \hat{v}_h \in V_h. \tag{5.1.21}$$

The unknown $\hat{u}_h^{n+1,k+1}$ has to vanish at 0 and L, thus it remains only the unknown values $u_i^{n+1,k+1}$ for $i = 1, \dots, N_x$. By choosing $\hat{v}_h = \hat{v}_h^j$ the function such that $\hat{v}_h^j(x_k) = 1$ if $k = j$ and 0 if $k \neq j$, it yields

$$\int_0^L \bar{\rho}_h^{n+1} \frac{\bar{u}_h^{n+1,k+1} - \bar{u}_h^{n+\frac{1}{2}}}{\Delta t} \bar{v}_h^j = \sum_{i=1}^{N_x} h \rho_i^{n+1/2} \frac{u_i^{n+1,k+1} - u_i^{n+\frac{1}{2}}}{\Delta t} v_i^j = h \rho_j^{n+1/2} \frac{u_j^{n+1,k+1} - u_j^{n+\frac{1}{2}}}{\Delta t}, \quad j = 1, \dots, N_x, \tag{5.1.22}$$

$$\int_0^L \frac{\partial_x \hat{u}_h^{n+1,k+1}}{\sqrt{|\partial_x \hat{u}_h^{n+1,k}|^2 + \varepsilon^2}} \partial_x \hat{v}_h^j = \begin{cases} \frac{u_1^{n+1,k+1} - u_2^{n+1,k+1}}{\sqrt{h^2\varepsilon^2 + \left|u_1^{n+1,k} - u_2^{n+1,k}\right|^2}} + \frac{u_1^{n+1,k+1}}{\sqrt{\frac{h^2}{4}\varepsilon^2 + (u_1^{n+1,k})^2}} & j = 1, \\ \frac{u_j^{n+1,k+1} - u_{j+1}^{n+1,k+1}}{\sqrt{h^2\varepsilon^2 + \left|u_{j+1}^{n+1,k} - u_j^{n+1,k}\right|^2}} + \frac{u_j^{n+1,k+1} - u_{j-1}^{n+1,k+1}}{\sqrt{h^2\varepsilon^2 + \left|u_j^{n+1,k} - u_{j-1}^{n+1,k}\right|^2}} & 2 \leq j \leq N_x - 1, \\ \frac{-u_{N_x-1}^{n+1,k+1} + u_{N_x}^{n+1,k+1}}{\sqrt{h^2\varepsilon^2 + \left|u_{N_x}^{n+1,k} - u_{N_x-1}^{n+1,k}\right|^2}} + \frac{u_{N_x}^{n+1,k+1}}{\sqrt{\frac{h^2}{4}\varepsilon^2 + (u_{N_x}^{n+1,k})^2}} & j = N_x, \end{cases} \tag{5.1.23}$$

$$\int_0^L \bar{f}\bar{v}_h^j = \begin{cases} h\left(\dfrac{f_0}{12} + \dfrac{f_1}{2} + \dfrac{f_2}{6}\right) & j = 1, \\[2ex] h\left(\dfrac{f_{j-1}}{6} + \dfrac{2f_j}{3} + \dfrac{f_{j+1}}{6}\right) & 2 \le j \le N_x - 1, \\[2ex] h\left(\dfrac{f_{N_x-1}}{6} + \dfrac{f_N}{2} + \dfrac{f_{N_x+1}}{12}\right) & j = N_x. \end{cases} \tag{5.1.24}$$

To find $u_h^{n+1,k+1}$, one has to solve the system of linear equations $AX = F$ as

$$\begin{pmatrix} a_1 & -b_1 & & & & \\ -b_1 & a_2 & -b_2 & & & \\ & -b_2 & a_3 & -b_3 & & \\ & & & \ddots & & \\ & & & -b_{N_x-2} & a_{N_x-1} & -b_{N_x-1} \\ & & & & -b_{N_x-1} & a_{N_x} \end{pmatrix} \begin{pmatrix} u_1^{n+1,k+1} \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ u_{N_x}^{n+1,k+1} \end{pmatrix} = \begin{pmatrix} F_1 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ F_{N_x} \end{pmatrix}, \tag{5.1.25}$$

where

$$b_i = \frac{\sigma_0 \Delta t}{\sqrt{h^2\varepsilon^2 + \left|u_{i+1}^{n+1,k} - u_i^{n+1,k}\right|^2}}, \qquad i = 1, \dots, N_x - 1,$$

$$\begin{aligned} a_1 &= \frac{\sigma_0 \Delta t}{\sqrt{h^2\varepsilon^2 + \left|u_2^{n+1,k} - u_1^{n+1,k}\right|^2}} + \frac{\sigma_0 \Delta t}{\sqrt{\frac{h^2}{4}\varepsilon^2 + \left(u_1^{n+1,k}\right)^2}} + h\rho_1^{n+1/2} \\ &= b_1 + \frac{\sigma_0 \Delta t}{\sqrt{\frac{h^2}{4}\varepsilon^2 + \left(u_1^{n+1,k}\right)^2}} + h\rho_1^{n+1/2}, \end{aligned}$$

$$\begin{aligned} a_{N_x} &= \frac{\sigma_0 \Delta t}{\sqrt{h^2\varepsilon^2 + \left|u_{N_x-1}^{n+1,k} - u_{N_x}^{n+1,k}\right|^2}} + \frac{\sigma_0 \Delta t}{\sqrt{\frac{h^2}{4}\varepsilon^2 + \left(u_{N_x}^{n+1,k}\right)^2}} + h\rho_{N_x}^{n+1/2} \\ &= b_{N_x-1} + \frac{\sigma_0 \Delta t}{\sqrt{\frac{h^2}{4}\varepsilon^2 + \left(u_{N_x}^{n+1,k}\right)^2}} + h\rho_{N_x}^{n+1/2}, \end{aligned}$$

$$\begin{aligned} a_i &= \frac{\sigma_0 \Delta t}{\sqrt{h^2\varepsilon^2 + \left|u_{i-1}^{n+1,k} - u_i^{n+1,k}\right|^2}} + \frac{\sigma_0 \Delta t}{\sqrt{h^2\varepsilon^2 + \left|u_i^{n+1,k} - u_{i+1}^{n+1,k}\right|^2}} + h\rho_i^{n+1/2} \\ &= b_{i-1} + b_i + h\rho_i^{n+1/2}, \qquad\qquad i = 2, \dots, N_x - 1, \end{aligned}$$

$$F_1 = h\Delta t\left(\frac{f_0}{12} + \frac{f_1}{2} + \frac{f_2}{6}\right) + h\rho_1^{n+1/2}u_1^n, \qquad F_{N_x} = h\Delta t\left(\frac{f_{N_x-1}}{6} + \frac{f_N}{2} + \frac{f_{N_x+1}}{12}\right) + h\rho_{N_x}^{n+1/2}u_{N_x}^n,$$

$$F_i = h\Delta t\left(\frac{f_{i-1}}{6} + \frac{2f_i}{3} + \frac{f_{i+1}}{6}\right) + h\rho_i^{n+1/2}u_i^n, \qquad i = 2, \dots, N_x - 1.$$

This tridiagonal system is resolved classically writing

$$u_i^{n+1,k+1} = \lambda_i u_{i+1}^{n+1,k+1} + r_i, \quad \text{for } i = 1, \dots, N_x - 1. \tag{5.1.26}$$

The system (5.1.25) indeed gives the recursive formulas

$$\lambda_1 = \frac{b_1}{a_1}, \qquad \lambda_i = \frac{b_i}{a_i - b_{i-1}\lambda_{i-1}} \quad \text{for } i = 2, \dots, N_x - 1, \tag{5.1.27}$$

$$r_1 = \frac{F_1}{a_1}, \qquad r_i = \frac{F_i + b_{i-1}r_{i-1}}{a_i - b_{i-1}\lambda_{i-1}} \quad \text{for } i = 2, \dots, N_x. \tag{5.1.28}$$

We find $u^{n+1,k+1}$ by setting $u_{N_x}^{n+1,k+1} = r_{N_x}$ and applying (5.1.26).

The previous procedure enables to find $u^{n+1,k+1}$ from the knowledge of $u^{n+1,k}$. We stop the iteration procedure when a stopping criterion is satisfied, $\|u^{n+1,k+1} - u^{n+1,k}\| \leq \varepsilon_{tol}$.

The numerical results are shown on Figure 5.2 with $N_x = 300$ and $\varepsilon_{tol} = 1\text{e-}7$. It uses 32603 iterations in time with the final time $T = 1$ and $\Delta t = 1.114\text{e-}5$ for the final time step. As can be

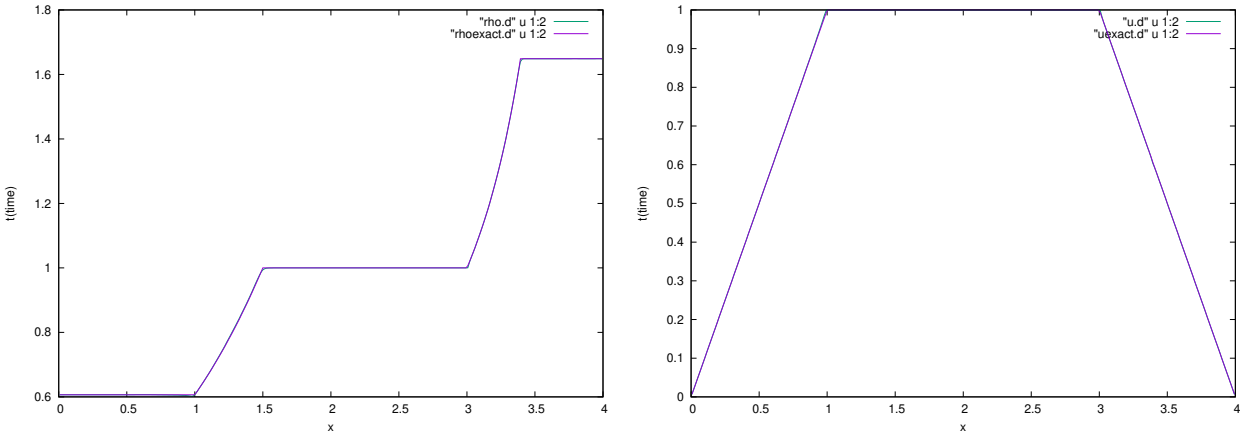


Figure 5.2: One-dimensional compressible Bingham model: comparison between exact and approximate solution computed by the regularisation method. Left: ρ, right: u.

seen in the left subfigure of Figure 5.2 there is no significant difference between the computed ρ (red line) and the exact ρ (green line) at the final time $T = 1$. We have the same conclusion for u. In fact, the L^1 error in this case is 3.872e-3 i.e. $\|(\hat{\rho}_h^{N_T} - \rho(T)\| + \|\hat{u}_h^{N_T} - u(T))\|_{L^1(0,L)}$=3.872e-3.

5.2 2D compressible Bingham model

5.2.1 Steady case

In this subsection the primal-dual algorithm is considered to solve the two-dimensional steady Bingham equation

$$\alpha u - \operatorname{div}\left(\sqrt{2}\frac{Du}{|Du|}\right) = f, \quad \text{for } (x,y) \in (-1,1) \times (-1,1). \tag{5.2.1}$$

An analytical solution is built under the form $u(x,y) = \Phi(r)\begin{pmatrix} -y \\ x \end{pmatrix}$ where $r = \sqrt{x^2+y^2}$. Then $|Du|/\sqrt{2} = r|\partial_r \Phi|/2$ and we have the equation

$$\left(\alpha\Phi(r) - \frac{\partial_r(\text{sgn } \partial_r\Phi)}{r} - \frac{2\text{sgn } \partial_r\Phi}{r^2}\right)\begin{pmatrix} -y \\ x \end{pmatrix} = f. \tag{5.2.2}$$

We consider
For $0 \le r \le 1/6$,

$$\text{sgn } \partial_r\Phi = (12r - 36r^2)^2, \quad \Phi = 1, \quad f = \left(\alpha - 2 \times 12^2(1-3r)(2-9r)\right)\begin{pmatrix} -y \\ x \end{pmatrix}.$$

For $1/6 \le r \le 1/3$,

$$\text{sgn } \partial_r\Phi = 1, \quad \Phi = 6r, \quad f = \left(6\alpha r - 2/r^2\right)\begin{pmatrix} -y \\ x \end{pmatrix}.$$

For $1/3 \le r \le 1/2$,

$$\text{sgn } \partial_r\Phi = \cos(\pi(6r-2)), \quad \Phi = 2, \quad f = \left(2\alpha + \frac{6\pi\sin(\pi(6r-2))}{r} - \frac{2\cos(\pi(6r-2))}{r^2}\right)\begin{pmatrix} -y \\ x \end{pmatrix}.$$

For $1/2 \le r \le 5/6$,

$$\operatorname{sgn} \partial_r \Phi = -1, \quad \Phi = 5 - 6r, \quad f = \big(\alpha(5-6r) + 2/r^2\big) \begin{pmatrix} -y \\ x \end{pmatrix}.$$

For $5/6 \le r \le 1$,

$$\operatorname{sgn} \partial_r \Phi = -\frac{1+\cos(\pi(6r-5))}{2}, \quad \Phi = 0, \quad f = \left(\frac{-3\pi \sin(\pi(6r-5))}{r} + \frac{1+\cos(\pi(6r-5))}{r^2} \right) \begin{pmatrix} -y \\ x \end{pmatrix}.$$

For $1 \le r$,

$$\operatorname{sgn} \partial_r \Phi = 0, \quad \Phi = 0, \quad f = 0.$$

Space discretisation: We consider a rectangular domain $\Omega = (-L_x, L_x) \times (-L_y, L_y)$, and we discretize it in both directions similarly as we did in the 1d case, see Figure 5.1. We denote by $N_x + 2$ and $N_y + 2$ the numbers of points in the horizontal and vertical directions. We define the spatial steps $h_x = \dfrac{2L_x}{N_x}$, $h_y = \dfrac{2L_y}{N_y}$ and the grid points as

$$x_0 = -L_x, \qquad x_{N_x+1} = L_x, \qquad y_0 = -L_y, \qquad y_{N_y+1} = L_y, \tag{5.2.3}$$

$$x_i = -L_x + \left(i - \frac{1}{2}\right) h_x, \qquad i = 1, \cdots, N_x, \tag{5.2.4}$$

$$y_i = -L_y + \left(j - \frac{1}{2}\right) h_y, \qquad j = 1, \cdots, N_y, \tag{5.2.5}$$

$$x_{i+\frac{1}{2}} = \frac{1}{2}(x_i + x_{i+1}), \qquad i = 0, \cdots, N_x, \tag{5.2.6}$$

$$y_{i+\frac{1}{2}} = \frac{1}{2}(y_i + y_{i+1}), \qquad j = 0, \cdots, N_y. \tag{5.2.7}$$

For each couple of integers i,j such that $0 \le i \le N_x + 1$, $0 \le j \le N_y + 1$, we define the rectangular cells as

$$R_{i,j} = (x_i, x_{i+1}) \times (y_j, y_{j+1}). \tag{5.2.8}$$

In order to get a mesh of triangles, these cells are cut in two along one diagonal, see Figure 5.3(left) giving the cells $K \in \mathcal{T}_h$.

We use the primal-dual algorithm, under the form (4.2.52) modified as (4.2.64) (mass lumped scheme). We have here $F(D) = \sqrt{2}|D|$, thus u_h^k and σ_h^k being known we look for u_h^{k+1} and σ_h^{k+1} solution to

$$\sigma_h^{k+1} = \mathbb{P}_r(\sigma_h^k + r(2D\hat{u}_h^k - D\hat{u}_h^{k-1})), \tag{5.2.9}$$

$$\alpha \int_\Omega \bar{u}_h^{k+1} \cdot \bar{v}_h \,\mathrm{dx} + \int_\Omega \sigma_h^{k+1} : D\hat{v}_h \,\mathrm{dx} + \int_\Omega \frac{\bar{u}_h^{k+1} - \bar{u}_h^k}{\tau} \cdot \bar{v}_h \,\mathrm{dx} = \int_\Omega \bar{f}_h \cdot \bar{v}_h \,\mathrm{dx} \quad \forall \hat{v}_h \in V_h, \tag{5.2.10}$$

where

$$\mathbb{P}_r(\sigma) = \begin{cases} \sigma & \text{if } |\sigma| \le \sqrt{2}, \\ \sqrt{2}\dfrac{\sigma}{|\sigma|} & \text{otherwise.} \end{cases} \tag{5.2.11}$$

We denote $u_h = (u_i)_{i\in I} = (u_i^x, u_i^y)_{i\in I}$ where I is a set of indices corresponding to the nodes in the domain. For $n \in I$, the finite volume cell around the node n is denoted by Q_n. Then one has

$$\alpha \int_\Omega \bar{u}_h^{k+1} \cdot \bar{v}_h \,\mathrm{dx} = \alpha \sum_{n\in I} |Q_n| u_n^{k+1} \cdot v_n = \alpha \sum_{n\in I} |Q_n| (u_n^{x,k+1} v_n^x + u_n^{y,k+1} v_n^y),$$

$$\int_\Omega \frac{\bar{u}_h^{k+1} - \bar{u}_h^k}{\tau} \cdot \bar{v}_h \,\mathrm{dx} = \sum_{n\in I} |Q_n| \frac{u_n^{k+1} - u_n^k}{\tau} \cdot v_n = \sum_{n\in I} |Q_n| \left(\frac{u_n^{x,k+1} - u_n^{x,k}}{\tau} v_n^x + \frac{u_n^{y,k+1} - u_n^{y,k}}{\tau} v_n^y \right),$$

$$\int_\Omega \bar{f}_h \cdot \bar{v}_h \,\mathrm{dx} = \sum_{n\in I} |Q_n| f_n \cdot v_n = \sum_{n\in I} |Q_n| (f_n^x v_n^x + f_n^y v_n^y).$$

Then, writing $\hat{v}_h$ in terms of basis function as $\hat{v}_h = \sum_{n\in I} v_n \varphi_n$ where v_n is a constant vector and where φ_n is an affine function such that $\varphi(x_n) = 1$ and $\varphi(x_m) = 0$ for all $m \neq n$, one has

$$D\hat{v}_h = \sum_n \frac{v_n \otimes \nabla\varphi_n + \nabla\varphi_n \otimes v_n}{2} = \sum_n \begin{pmatrix} v_n^x \partial_x \varphi_n & \frac{v_n^x \partial_y \varphi_n + v_n^y \partial_x \varphi_n}{2} \\ \frac{v_n^x \partial_y \varphi_n + v_n^y \partial_x \varphi_n}{2} & v_n^y \partial_y \varphi_n \end{pmatrix}.$$

It gives

$$\int_\Omega \sigma : D\hat{v}_h = \sum_{K\in\mathcal{T}_h} |K| \sigma_K : (D\hat{v}_h)_K \tag{5.2.12}$$

$$= \sum_{K\in\mathcal{T}_h} \sum_{n\in I} |K| \Big(\sigma_{K,xx} v_n^x (\partial_x \varphi_n)_{|K} + \sigma_{K,yy} v_n^y (\partial_y \varphi_n)_{|K} + \sigma_{K,xy} \big(v_n^x (\partial_y \varphi_n)_{|K} + v_n^y (\partial_x \varphi_n)_{|K} \big) \Big). \tag{5.2.13}$$

By choosing $\hat{v}_h = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \varphi_n$ in (5.2.10) we get for $n \in I$

$$\alpha |Q_n| u_n^{x,k+1} + |Q_n| \frac{u_n^{x,k+1} - u_n^{x,k}}{\tau} + \sum_{K\in V(n)} |K| \big(\sigma_{K,xx}^{k+1} (\partial_x \varphi_n)_{|K} + \sigma_{K,xy}^{k+1} (\partial_y \varphi_n)_{|K} \big) = |Q_n| f_n^x. \tag{5.2.14}$$

Similarly by choosing $\hat{v}_h = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \varphi_n$ it gives for $n \in I$

$$\alpha |Q_n| u_n^{y,k+1} + |Q_n| \frac{u_n^{y,k+1} - u_n^{y,k}}{\tau} + \sum_{K\in V(n)} |K| \big(\sigma_{K,yy}^{k+1} (\partial_y \varphi_n)_{|K} + \sigma_{K,xy}^{k+1} (\partial_x \varphi_n)_{|K} \big) = |Q_n| f_n^y. \tag{5.2.15}$$

Therefore we can compute explicitly u_h^{k+1} as

$$\begin{cases} u_n^{x,k+1} = \frac{1}{\alpha + \frac{1}{\tau}} \left(f_n^x + \frac{1}{\tau} u_n^{x,k} - \sum_{K\in V(n)} \frac{|K|}{|Q_n|} \big(\sigma_{K,xx}^{k+1} (\partial_x \varphi_n)_{|K} + \sigma_{K,xy}^{k+1} (\partial_y \varphi_n)_{|K} \big) \right), \\ u_n^{y,k+1} = \frac{1}{\alpha + \frac{1}{\tau}} \left(f_n^y + \frac{1}{\tau} u_n^{y,k} - \sum_{K\in V(n)} \frac{|K|}{|Q_n|} \big(\sigma_{K,yy}^{k+1} (\partial_y \varphi_n)_{|K} + \sigma_{K,xy}^{k+1} (\partial_x \varphi_n)_{|K} \big) \right). \end{cases} \tag{5.2.16}$$

An illustration is on Figure 5.3(right).

Primal-dual algorithm on a structured mesh using Fortran. To avoid any geometric complexity we use the above structured mesh together with Fortran coding for the primal-dual algorithm (without acceleration). The results are shown on Figure 5.4.

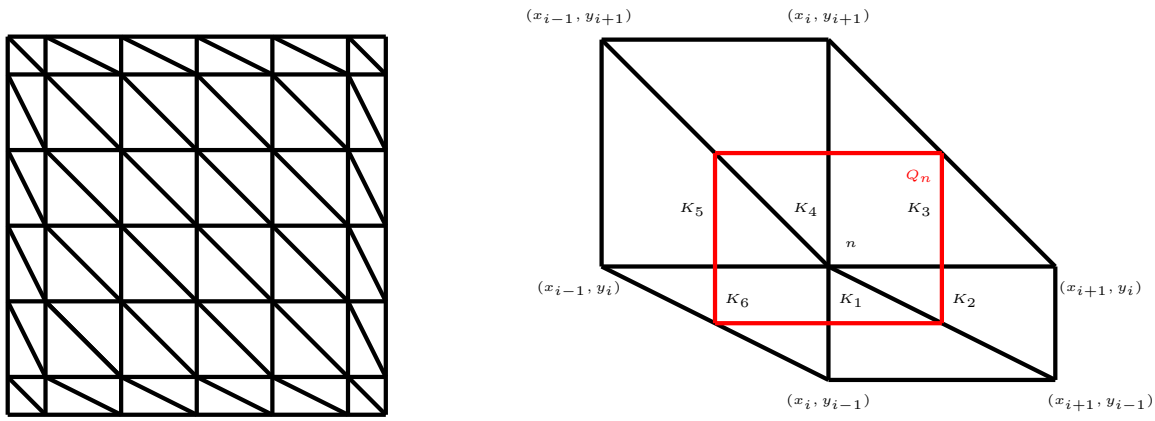


Figure 5.3: Space discretisation

$r\tau$	Number of iterations	Error
$> h^2/6$	NaN	NaN
$h^2/6$	4103	1.0683
$h^2/8$	4407	0.1068
$h^2/16$	5076	0.1067
$h^2/32$	6270	0.1067

Table 5.1: Test with various values of $r\tau$

Optimal constant C in the formula $r\tau = Ch^2$

In the convergence of the primal-dual algorithm, Theorem 4.2.6 requires the stability condition $r\tau L_h^2 \leq 1$ where L_h is defined according to (4.2.65) by

$$\|D\hat{v}_h\|_{L^2} \leq L_h \left(\int_\Omega |\bar{v}_h|^2 \right)^{1/2}, \quad \forall \hat{v}_h \in V_h. \tag{5.2.17}$$

Since obviously L_h is of the order of $1/h$, the stability condition means that $r\tau \leq Ch^2$ for some appropriate constant C. We test various values of the product $r\tau$ in the case $h_x = h_y = 1/20$ with the stopping criteria $\varepsilon_{tol} = 10^{-6}$ and $r = \tau = \sqrt{C}h$, $\alpha = 1$. As can be seen in Table 5.1, the error increases as C decreases, even though it decreases significantly in the case $C = 1/8$ in comparison with $C = 1/6$, and then decreases slowly with smaller values of C. On the other hand, the number of iteration increases proportionally to the increase of $1/C$. Besides, the algorithm does not converge for values higher than $1/6$. Thus we choose $C = 1/8$ as the optimal value for $h = 1/20$.

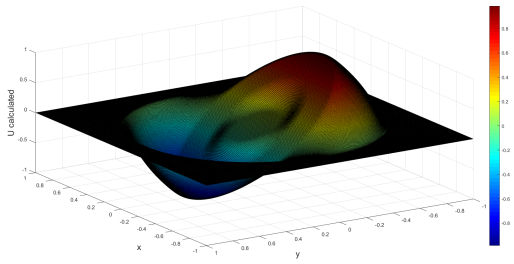
Testing with $h = 1/20, 1/40, 1/80$ the optimal C is found to be $1/6$ with respect to both the number of iterations and the error.

Optimal choice for the stopping criterion ε_{tol} for the primal-dual algorithm

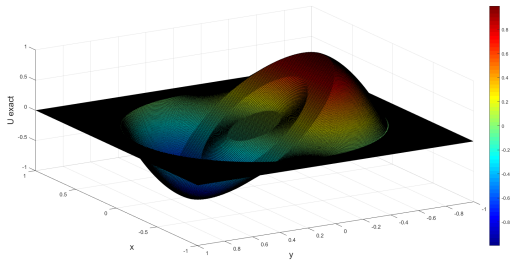
In the first primal-dual algorithm we have set a stopping criterion $\|\hat{u}_h^{k+1} - \hat{u}_h^k\|_{L^2} \leq \varepsilon_{tol}$ for the iteration loop. A smaller ε_{tol} means a smaller error, but more iterations. According to Table 5.2 showing the error, we can find the optimal stopping value for each value of h. In this test we fix $\alpha = 1$, $r = \tau = \frac{h}{\sqrt{8}}$.

Optimal choice of the parameters r, τ for the primal-dual algorithm

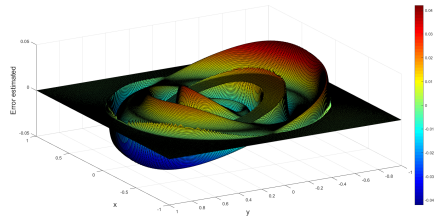
Starting from the term $\sigma + rDu$, we expect that $r \sim \frac{[\sigma]}{[Du]} \sim h\frac{[\sigma]}{[u]}$, where the square brackets represent



(a) u_h^x computed



(b) u_h^x exact



(c) Error between u_h^x exact and u_h^x computed. It is everywhere less than 4%.

Figure 5.4: Primal-dual algorithm with structured mesh for $h = 1/80$

ε_{tol}	$h = 1/20$	$h = 1/40$	$h = 1/80$
10^{-2}	0.1713	0.4452	0.1713
10^{-3}	**9.5816e-2**	8.2301e-2	8.5686e-2
10^{-4}	0.1038	**7.7183e-2**	4.5593e-2
10^{-5}	0.1054	7.9293e-2	**4.5555e-2**
10^{-6}	0.1064	7.9830e-2	4.6098e-2

Table 5.2: Computed error for different values of h and ε_{tol}. Boldface numbers correspond to the optimal choice of ε_{tol}, the largest for which we obtain the floor value of the error.

the order of magnitude of a quantity. Hence the expected order of magnitude of r is

$$r \sim h\frac{[\sigma]}{[u]}. \tag{5.2.18}$$

In the iteration scheme one can expect that the term αu has the same order of magnitude as the penalty term $\frac{u^{k+1}-u^k}{\tau}$. Thus τ is comparable to $\frac{1}{\alpha}$. On the other hand, from the equation $\alpha u - \operatorname{div}\sigma = f$, one has $\alpha \sim \frac{[\operatorname{div}\sigma]}{[u]} \sim \frac{[\sigma]}{h[u]}$. Hence an expected order of magnitude of τ is

$$\tau \sim \frac{h[u]}{[\sigma]}. \tag{5.2.19}$$

We notice that the quantity $[u]$ in (5.2.18) and (5.2.19) can eventually be replaced by $\frac{[f]}{\alpha}$, because from the equation $\alpha u - \operatorname{div}\sigma = f$ one has merely $[f] \sim \alpha[u]$. For a time-dependent problem one can use instead $[u] = [u_0] + t[f]$, where u_0 is the initial datum.

In Tables 5.3, 5.4, 5.5, we give the error and the cost for several choices of r and τ and different values of h. The optimal value in the three tables is $r = \frac{h}{16}$ and $\tau = 2h$. Notice that the choice for

(τ, r)	$\left(\frac{h^2}{8}, 1\right)$	$\left(\frac{h}{16}, 2h\right)$	$\left(\frac{h}{8}, h\right)$	$\left(\frac{h}{4}, \frac{h}{2}\right)$	$\left(\frac{h}{\sqrt{8}}, \frac{h}{\sqrt{8}}\right)$	$\left(\frac{h}{2}, \frac{h}{4}\right)$	$\left(h, \frac{h}{8}\right)$	$\left(2h, \frac{h}{16}\right)$	$\left(1, \frac{h^2}{8}\right)$
N° of iterations	30	272	270	236	236	233	193	**147**	908
Error	0.9375	0.3424	0.1443	9.7019e-2	9.5816e-2	9.8947e-2	9.4793e-2	**6.7218e-2**	0.1649

Table 5.3: Number of iterations and error for different choices of τ, r for $h = 1/20$, $\varepsilon_{tol} = 10^{-3}$.

(τ, r)	$\left(\frac{h^2}{8}, 1\right)$	$\left(\frac{h}{16}, 2h\right)$	$\left(\frac{h}{8}, h\right)$	$\left(\frac{h}{4}, \frac{h}{2}\right)$	$\left(\frac{h}{\sqrt{8}}, \frac{h}{\sqrt{8}}\right)$	$\left(\frac{h}{2}, \frac{h}{4}\right)$	$\left(h, \frac{h}{8}\right)$	$\left(2h, \frac{h}{16}\right)$	$\left(1, \frac{h^2}{8}\right)$
N° of iterations	242	1559	1002	849	775	695	489	**370**	4174
Error	0.9402	0.1044	7.6578e-2	7.6836e-2	7.7183e-2	7.6476e-2	6.8291e-2	**4.8981e-2**	7.0575e-2

Table 5.4: Number of iterations and error for different choices of τ, r for $h = 1/40$, $\varepsilon_{tol} = 10^{-4}$.

which the error is the smallest is also the choice for which the number of iterations is the smallest. Considering that here $[u] = 2$, the previous considerations lead to the "optimal" choice

$$r = \frac{1}{8} h \frac{[\sigma]}{[u]}, \qquad \tau = \frac{h[u]}{[\sigma]}. \tag{5.2.20}$$

If we replace $[u]$ by $\frac{[f]}{\alpha}$ and considering that here $\frac{[f]}{\alpha} = 32$, this leads rather to the formula

$$r = 2h \frac{\alpha[\sigma]}{[f]}, \qquad \tau = \frac{1}{16} \frac{h[f]}{\alpha[\sigma]}. \tag{5.2.21}$$

Primal-dual algorithm without mass lumping

Consider now using the original primal-dual iteration method (4.2.52) without the mass lumping modification (4.2.64). In other words we use the FEM scalar product instead of the FVM scalar product. The main difference between both methods is that now one has to solve a linear system of equations. Its cost is known to be generically $O(n^2 \log n)$ for a general matrix, and $O(n \log n)$ in case of a "good" sparse matrix, with n the size of the matrix.

For dealing with the method without mass lumping we use the software FreeFEM++, that automatically generates a mesh. For it we choose the number of intervals $n = 40$, we take $u_{init} = \sigma_{init} = 0$, $\tau = r = h/6$. We make the test for a scalar function (i.e. for the so called TV minimization problem), with various values of ε_{tol}. The results are as follows.

- Without mass lumping
  $\varepsilon = 10^{-2}$ err $= 0.332847$ n.o.iter $= 153$ execution time $= 26.4396s$
  $\varepsilon = 10^{-3}$ err $= 0.113881$ n.o.iter $= 418$ execution time $= 69.0468s$
  $\varepsilon = 10^{-4}$ err $= 0.10936$ n.o.iter $= 746$ execution time $= 112.164s$
  $\varepsilon = 10^{-5}$ err $= 0.109946$ n.o.iter $= 1055$ execution time $= 172.84s$

(τ, r)	$\left(\frac{h^2}{8}, 1\right)$	$\left(\frac{h}{16}, 2h\right)$	$\left(\frac{h}{8}, h\right)$	$\left(\frac{h}{4}, \frac{h}{2}\right)$	$\left(\frac{h}{\sqrt{8}}, \frac{h}{\sqrt{8}}\right)$	$\left(\frac{h}{2}, \frac{h}{4}\right)$	$\left(h, \frac{h}{8}\right)$	$\left(2h, \frac{h}{16}\right)$	$\left(1, \frac{h^2}{8}\right)$
N° of iterations	31915	5174	3218	2466	2162	1858	1283	**1153**	18882
Error	0.5222	4.9833e-2	4.6645e-2	4.6131e-2	4.5556e-2	4.4453e-2	4.0351e-2	**3.2291e-2**	3.3465e-2

Table 5.5: Number of iterations and error for different choices of τ, r for $h = 1/80$, $\varepsilon_{tol} = 10^{-5}$.

(τ_0, r_0)	$\left(\frac{h^2}{8}, 1\right)$	$(\frac{h}{16}, 2h)$	$(\frac{h}{8}, h)$	$(\frac{h}{4}, \frac{h}{2})$	$\left(\frac{h}{\sqrt{8}}, \frac{h}{\sqrt{8}}\right)$	$(\frac{h}{2}, \frac{h}{4})$	$(h, \frac{h}{8})$	$(2h, \frac{h}{16})$	$\left(1, \frac{h^2}{8}\right)$
N° of iterations	232	106	86	69	69	**69**	70	71	73
Error	0.3931	0.1130	6.5399e-2	4.1523e-2	3.6311e-2	**3.6448e-2**	4.0468e-2	4.3254e-2	4.4897e-2

Table 5.6: Number of iterations and error for the Accelerated scheme with different choices of τ_0, r_0 for $h = 1/20$, $\varepsilon_{tol} = 10^{-3}$.

(τ_0, r_0)	$\left(\frac{h^2}{8}, 1\right)$	$(\frac{h}{16}, 2h)$	$(\frac{h}{8}, h)$	$(\frac{h}{4}, \frac{h}{2})$	$\left(\frac{h}{\sqrt{8}}, \frac{h}{\sqrt{8}}\right)$	$(\frac{h}{2}, \frac{h}{4})$	$(h, \frac{h}{8})$	$(2h, \frac{h}{16})$	$\left(1, \frac{h^2}{8}\right)$
N° of iterations	1843	517	484	365	365	**365**	367	368	370
Error	0.2497	5.1105e-2	2.9266e-2	1.8845e-2	1.6860e-2	**1.6701e-2**	1.7926e-2	1.8984e-2	1.9894e-2

Table 5.7: Number of iterations and error for the Accelerated scheme with different choices of τ_0, r_0 for $h = 1/40$, $\varepsilon_{tol} = 10^{-4}$.

- With mass lumping
  $\varepsilon = 10^{-2}$ err $= 0.770504$ n.o.iter $= 92$ execution time $= 12.4226s$
  $\varepsilon = 10^{-3}$ err $= 0.112977$ n.o.iter $= 419$ execution time $= 52.3369s$
  $\varepsilon = 10^{-4}$ err $= 0.111199$ n.o.iter $= 752$ execution time $= 93.9013s$
  $\varepsilon = 10^{-5}$ err $= 0.111853$ n.o.iter $= 1077$ execution time $= 134.989s$

In conclusion, the primal-dual algorithm without or with mass lumping give the same quality of approximation for the same number of iterations. However the mass-lumped method runs faster because there is no linear system to solve. The gain is not big here because the mass matrix is almost diagonal and it is not much costly to invert it.

5.2.2 Accelerated scheme

We evaluate here the accelerated primal-dual algorithm (4.2.61) (with mass lumping). For this method we have to choose r_0 and τ_0.

Optimal choice for r_0, τ_0

Similarly as in the previous non-accelerated scheme we try several values. The error and number of iterations are reported in Tables 5.6, 5.7, 5.8. We take here $\alpha = 20$. We can see that the most efficient choice is $(\tau_0, r_0) = (\frac{h}{2}, \frac{h}{4})$.

Comparison between the two methods

We choose $\alpha = 100$ and consider the normal and accelerated primal-dual algorithms. Following Tables 5.9, 5.10, 5.11, 5.12, for any value of h the accelerated scheme with the best parameters $(\tau_0, r_0) = (\frac{h}{2}, \frac{h}{4})$ gives only a slightly smaller error than the normal scheme with the best parameters $(\tau, r) = (\frac{h}{16}, 2h)$. Meanwhile, the accelerated scheme uses more iterations to attain this value of the

(τ_0, r_0)	$\left(\frac{h^2}{8}, 1\right)$	$(\frac{h}{16}, 2h)$	$(\frac{h}{8}, h)$	$(\frac{h}{4}, \frac{h}{2})$	$\left(\frac{h}{\sqrt{8}}, \frac{h}{\sqrt{8}}\right)$	$(\frac{h}{2}, \frac{h}{4})$	$(h, \frac{h}{8})$	$(2h, \frac{h}{16})$	$\left(1, \frac{h^2}{8}\right)$
N° of iterations	13172	2284	1812	1660	1659	**1660**	1661	1663	1666
Error	0.1581	2.4531e-2	1.5232e-2	1.0248e-2	9.5125e-3	**9.4371e-3**	9.8587e-3	1.0252e-2	1.0655e-2

Table 5.8: Number of iterations and error for the Accelerated scheme with different choices of τ_0, r_0 for $h = 1/80$, $\varepsilon_{tol} = 10^{-5}$.

(τ, r)	$\left(\frac{h^2}{8}, 1\right)$	$\left(\frac{h}{16}, 2h\right)$	$\left(\frac{h}{8}, h\right)$	$\left(\frac{h}{4}, \frac{h}{2}\right)$	$\left(\frac{h}{\sqrt{8}}, \frac{h}{\sqrt{8}}\right)$	$\left(\frac{h}{2}, \frac{h}{4}\right)$	$\left(h, \frac{h}{8}\right)$	$\left(2h, \frac{h}{16}\right)$	$\left(1, \frac{h^2}{8}\right)$
N° of iterations	111	**27**	32	49	60	73	103	4	3
Error	3.4703e-2	**1.9001e-2**	2.2308e-2	2.5919e-2	2.9566e-2	3.5135e-2	5.4811e-2	0.1299	0.1324
(τ_0, r_0)	$\left(\frac{h^2}{8}, 1\right)$	$\left(\frac{h}{16}, 2h\right)$	$\left(\frac{h}{8}, h\right)$	$\left(\frac{h}{4}, \frac{h}{2}\right)$	$\left(\frac{h}{\sqrt{8}}, \frac{h}{\sqrt{8}}\right)$	$\left(\frac{h}{2}, \frac{h}{4}\right)$	$\left(h, \frac{h}{8}\right)$	$\left(2h, \frac{h}{16}\right)$	$\left(1, \frac{h^2}{8}\right)$
N° of iterations	144	52	37	33	32	**32**	33	33	34
Error	0.1750	0.1130	4.9285e-2	1.9822e-2	1.8080e-2	**1.7683e-2**	1.7864e-2	1.8335e-2	1.8452e-2

Table 5.9: Number of iterations and error for normal and accelerated schemes with different choices of τ, r for $h = 1/20$, $\varepsilon_{tol} = 10^{-3}$, $\alpha = 100$

(τ, r)	$\left(\frac{h^2}{8}, 1\right)$	$\left(\frac{h}{16}, 2h\right)$	$\left(\frac{h}{8}, h\right)$	$\left(\frac{h}{4}, \frac{h}{2}\right)$	$\left(\frac{h}{\sqrt{8}}, \frac{h}{\sqrt{8}}\right)$	$\left(\frac{h}{2}, \frac{h}{4}\right)$	$\left(h, \frac{h}{8}\right)$	$\left(2h, \frac{h}{16}\right)$	$\left(1, \frac{h^2}{8}\right)$
N° of iterations	932	**99**	120	161	195	243	406	693	3
Error	2.7232e-2	**1.0971e-2**	1.3108e-2	1.6009e-2	1.7693e-2	1.9275e-2	2.2429e-2	2.6756e-2	0.1324
(τ_0, r_0)	$\left(\frac{h^2}{8}, 1\right)$	$\left(\frac{h}{16}, 2h\right)$	$\left(\frac{h}{8}, h\right)$	$\left(\frac{h}{4}, \frac{h}{2}\right)$	$\left(\frac{h}{\sqrt{8}}, \frac{h}{\sqrt{8}}\right)$	$\left(\frac{h}{2}, \frac{h}{4}\right)$	$\left(h, \frac{h}{8}\right)$	$\left(2h, \frac{h}{16}\right)$	$\left(1, \frac{h^2}{8}\right)$
N° of iterations	985	244	192	166	164	**164**	165	166	167
Error	0.1116	2.3446e-2	1.4898e-2	1.0437e-2	9.5771e-3	**9.3094e-3**	9.3612e-3	9.4723e-3	9.5632e-3

Table 5.10: Number of iterations and error for the normal and accelerated schemes with different choices of τ, r for $h = 1/40$, $\varepsilon_{tol} = 10^{-4}$, $\alpha = 100$

error, and this is worse for smaller h. However if we choose $(\tau, r) = (\frac{h}{\sqrt{8}}, \frac{h}{\sqrt{8}})$ for both methods, the accelerated scheme performs better. We conclude that finding optimal values for r, τ (or r_0, τ_0) is more important that choosing the normal or accelerated scheme. For the normal scheme the formula (5.2.21) performs quite well since $[f]/\alpha = 2$ here.

5.2.3 Implementation of various types of boundary conditions

In the previous tests we always implemented the Dirichlet boundary condition when solving the viscoplastic part (1.3.2). It is however possible to formulate and implement the finite element formulation in the case of other boundary conditions, by adding degrees of freedom on the boundary of the domain. One can treat in particular Neumann or slip boundary conditions. The implementation of a friction condition is a bit more difficult.

We consider the viscoplastic model

$$\alpha u - \operatorname{div} \sigma = f, \tag{5.2.22}$$

(τ, r)	$\left(\frac{h^2}{8}, 1\right)$	$\left(\frac{h}{16}, 2h\right)$	$\left(\frac{h}{8}, h\right)$	$\left(\frac{h}{4}, \frac{h}{2}\right)$	$\left(\frac{h}{\sqrt{8}}, \frac{h}{\sqrt{8}}\right)$	$\left(\frac{h}{2}, \frac{h}{4}\right)$	$\left(h, \frac{h}{8}\right)$	$\left(2h, \frac{h}{16}\right)$	$\left(1, \frac{h^2}{8}\right)$
N° of iterations	2688	**386**	529	739	886	1026	1385	2060	4
Error	7.3377e-3	**5.7931e-3**	6.4370e-3	7.6133e-3	8.4351e-3	9.7475e-3	1.3268e-2	1.6938e-2	0.1324
(τ_0, r_0)	$\left(\frac{h^2}{8}, 1\right)$	$\left(\frac{h}{16}, 2h\right)$	$\left(\frac{h}{8}, h\right)$	$\left(\frac{h}{4}, \frac{h}{2}\right)$	$\left(\frac{h}{\sqrt{8}}, \frac{h}{\sqrt{8}}\right)$	$\left(\frac{h}{2}, \frac{h}{4}\right)$	$\left(h, \frac{h}{8}\right)$	$\left(2h, \frac{h}{16}\right)$	$\left(1, \frac{h^2}{8}\right)$
N° of iterations	6531	1149	875	864	863	**862**	863	864	865
Error	7.0693e-2	1.0937e-2	7.7052e-3	5.6862e-3	5.3788e-3	**5.2727e-3**	5.2671e-3	5.3037e-3	5.3370e-3

Table 5.11: Number of iterations and error for the normal and accelerated schemes with different choices of τ, r for $h = 1/80$, $\varepsilon_{tol} = 10^{-5}$, $\alpha = 100$

(τ, r)	$\left(\frac{h^2}{8}, 1\right)$	$\left(\frac{h}{16}, 2h\right)$	$\left(\frac{h}{8}, h\right)$	$\left(\frac{h}{4}, \frac{h}{2}\right)$	$\left(\frac{h}{\sqrt{8}}, \frac{h}{\sqrt{8}}\right)$	$\left(\frac{h}{2}, \frac{h}{4}\right)$	$\left(h, \frac{h}{8}\right)$	$\left(2h, \frac{h}{16}\right)$	$\left(1, \frac{h^2}{8}\right)$
N° of iterations	12621	**1319**	2074	3078	3717	4479	6532	9264	4
Error	3.7658e-3	**3.2884e-3**	3.3942e-3	3.6961e-3	3.9771e-3	4.3723e-3	5.5668e-3	7.6527e-3	0.1325
(τ_0, r_0)	$\left(\frac{h^2}{8}, 1\right)$	$\left(\frac{h}{16}, 2h\right)$	$\left(\frac{h}{8}, h\right)$	$\left(\frac{h}{4}, \frac{h}{2}\right)$	$\left(\frac{h}{\sqrt{8}}, \frac{h}{\sqrt{8}}\right)$	$\left(\frac{h}{2}, \frac{h}{4}\right)$	$\left(h, \frac{h}{8}\right)$	$\left(2h, \frac{h}{16}\right)$	$\left(1, \frac{h^2}{8}\right)$
N° of iterations	58940	5104	3894	3582	3578	**3577**	3578	3579	3581
Error	6.3156e-2	5.4191e-3	4.0941e-3	3.3500e-3	3.2154e-3	**3.1653e-3**	3.1567e-3	3.1718e-3	3.1893e-3

Table 5.12: Number of iterations and error for the normal and accelerated schemes with different choices of τ, r for $h = 1/160$, $\varepsilon_{tol} = 10^{-6}$, $\alpha = 100$

$$\sigma \in \partial F(Du). \tag{5.2.23}$$

We have the corresponding variational formulation

$$\alpha \int u \cdot (v - u) + \int F(Dv) \geq \int F(Du) + \int f \cdot (v - u), \qquad \forall v \in V. \tag{5.2.24}$$

For Dirichlet boundary conditions we take

$$V = K_0^F, \tag{5.2.25}$$

see Definition 3.2.2.
For Neumann boundary conditions $\sigma n = 0$, we take

$$V = L^2. \tag{5.2.26}$$

For slip boundary conditions $u \cdot n = 0, (\sigma n) \times n = 0$, we take

$$V = \{u \in K^F \mid u \cdot n = 0\}. \tag{5.2.27}$$

Dirichlet condition. We use the radial solution

$$u(x, y) = \Phi(r) \begin{pmatrix} -y \\ x \end{pmatrix}, \tag{5.2.28}$$

where $r = \sqrt{x^2 + y^2}$, $\Omega = (-1, 1) \times (-1, 1)$, that has been used in subsection 5.2.1.

Neumann condition. We use the radial solution

$$u(x, y) = \Phi(r) \begin{pmatrix} -y \\ x \end{pmatrix} + \begin{pmatrix} 3 \\ 5 \end{pmatrix}, \tag{5.2.29}$$

which is a solution to (5.2.22) with $\alpha = 1$ and with Neumann condition if we add $\begin{pmatrix} 3 \\ 5 \end{pmatrix}$ to the previous f. We use 161 points in each direction. The stopping criterion is 10^{-6}.
The numerical results give the error $4.63.10^{-2}$, and the number of iterations 3740.
Note: We have to change the value of τ, r as $\tau = r = \dfrac{dx}{5}$. The results are shown on Figure 5.5.

Slip condition: $u \cdot n = 0$ and $(\sigma n) \times n = 0$. We take

$$u = \begin{pmatrix} u_x(x) \\ 0 \end{pmatrix}. \tag{5.2.30}$$

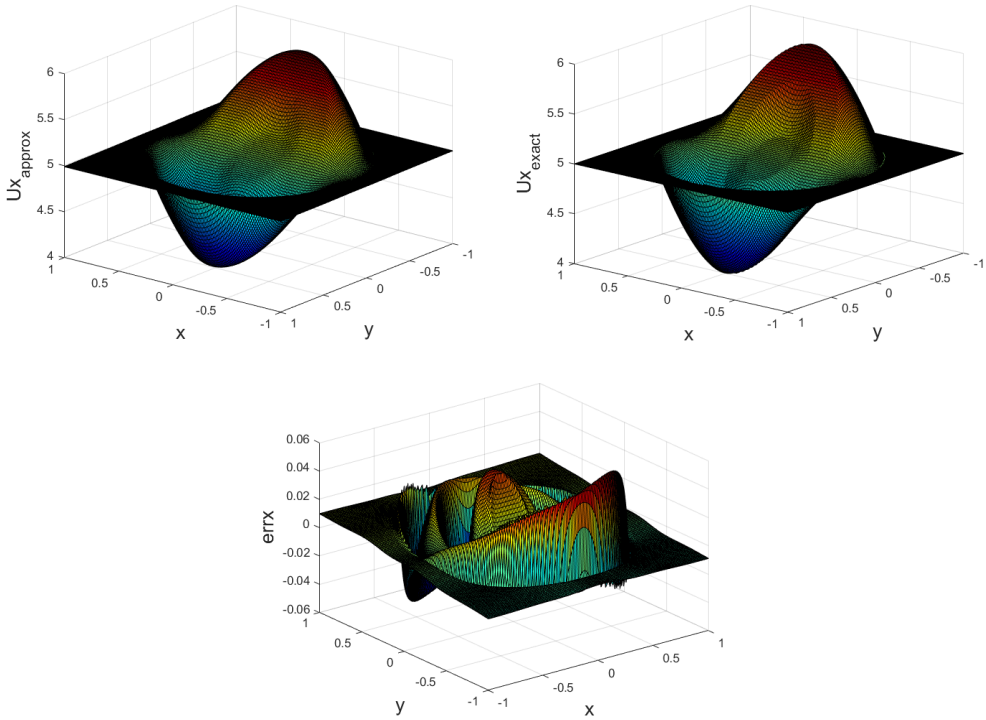


Figure 5.5: Numerical results for Neumann boundary condition

Then

$$Du = \begin{pmatrix} \partial_x u_x & 0 \\ 0 & 0 \end{pmatrix}, \tag{5.2.31}$$

$$\sigma = \begin{pmatrix} \mathrm{sgn}(\partial_x u_x) & 0 \\ 0 & 0 \end{pmatrix}, \tag{5.2.32}$$

for $F(Du) = |Du|$. On the right edge we have $n = (1,0)^T$, thus the boundary condition writes

$$u_x \mid_{x=1} = 0, \tag{5.2.33}$$

$$\begin{pmatrix} \mathrm{sgn}(\partial_x u_x) \\ 0 \end{pmatrix} \times \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 0. \tag{5.2.34}$$

This last condition holds trivially. Similarly we get $u_x \mid_{x=-1} = 0$. We take

$$u_x = \begin{cases} 2x+2 & \text{if } -1 \le x \le -0.5, \\ 1 & \text{if } -0.5 \le x \le 0.5, \\ -2x+2 & \text{if } 0.5 \le x \le 1, \end{cases} \tag{5.2.35}$$

$$\mathrm{sgn}(\partial_x u_x) = \begin{cases} 1 & \text{if } -1 \le x \le -0.5, \\ -\sin(\pi x) & \text{if } -0.5 \le x \le 0.5, \\ -1 & \text{if } 0.5 \le x \le 1, \end{cases} \tag{5.2.36}$$

$$f_x = \alpha u_x - \partial_x \mathrm{sgn}(\partial_x u_x) = \begin{cases} 2x+2 & \text{if } -1 \le x \le -0.5, \\ 1 + \pi\cos(\pi x) & \text{if } -0.5 \le x \le 0.5, \\ -2x+2 & \text{if } 0.5 \le x \le 1. \end{cases} \tag{5.2.37}$$

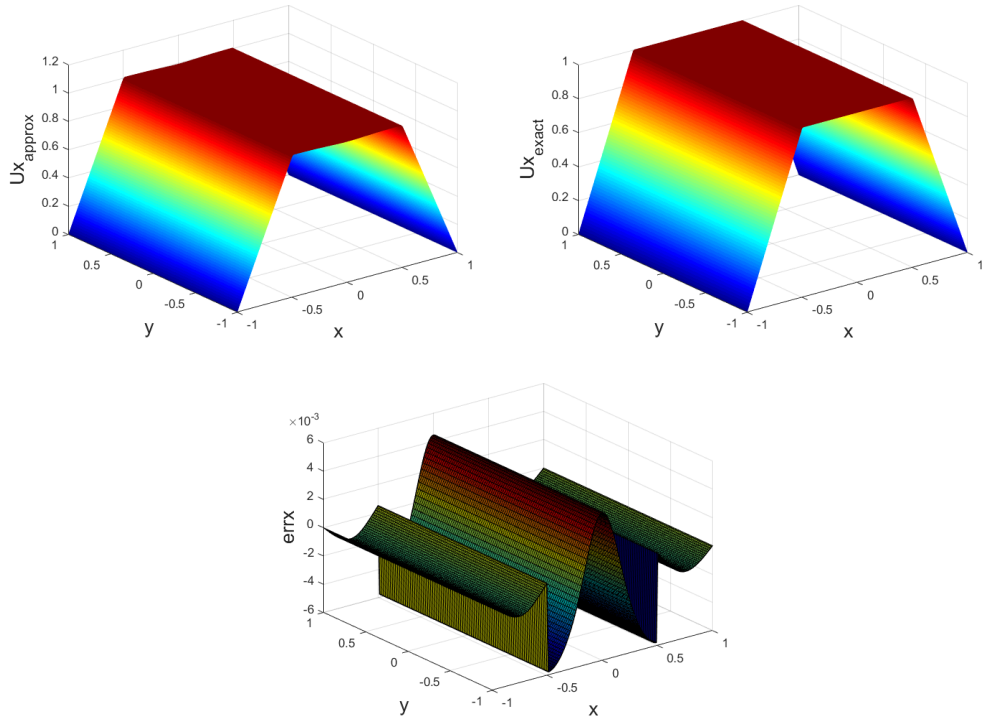


Figure 5.6: Numerical results for slip boundary condition

Note: We do not take any degree of freedom at the corners, do not take any degree of freedom for u_x at the right, left edges; do not take any degree of freedom for u_y at the top, bottom edges.
we take 161 points in each direction. The stopping criterion is 10^{-6}. We obtain
Error $5.55.10^{-3}$
Number of iterations 2071
The results are shown on Figure 5.6.

5.2.4 Comparison with the regularization and augmented Lagrangian methods

We implement the primal-dual algorithm, its accelerated version, the regularization method, and the augmented Lagrangian method within the same FreeFEM++ software, so as to compare them in particular concerning their execution time. The test case is the radial one of Subsection 5.2.1, with $\alpha = 100$. The results are shown in Tables 5.13, 5.14, 5.15, 5.16. We observe that for a given space resolution, the error that we obtain with each method (taking its best parameters) is more or less the same, the accelerated primal-dual method having a slightly smaller error. On the contrary the number of iterations varies a lot. More iterations are necessary with the primal-dual algorithm, and in particular with the accelerated algorithm (which is unexpected since it is supposed to converge faster). We observe that comparing the unstructured mesh used here with the Cartesian mesh used with the Fortran code, the primal-dual algorithm takes more or less the same number of iterations, whereas the accelerated version takes more iterations in the unstructured mesh configuration than in the Cartesian one (compare the last columns of Tables 5.13, 5.14 to Table 5.11). Then we have to compare execution time, which is the real issue since each step of the iteration does not have the same cost for all the methods (regularisation and augmented Lagrangian methods have linear systems to solve). We can see that the primal-dual algorithm takes more time

than the regularisation method, but is faster than the augmented Lagrangian method. In fact the augmented Lagrangian method can take more iterations. This maybe due to the fact that the penalty parameter r is not chosen optimally, indeed this is difficult to know the best value. We conclude that for this test the regularisation method is the best, knowing that in FreeFEM++ linear systems are solved very efficiently. Moreover here $\alpha = 100$ thus the problem is not stiff. We have to remember also that in some cases the regularisation method has difficulties to well localise the solid zones, what the primal-dual algorithm is supposed to do well.

$N_x = N_y$	20	40	80	160
Number of iterations	8	26	99	390
Execution time	0.4183	3.54173	45.1022	678.248
Error	0.06157	0.02445	0.00971	0.00424

Table 5.13: Error, number of iterations and execution time for the primal-dual algorithm for various spatial steps with optimal values $(\tau, r) = (h/16, 2h)$.

$N_x = N_y$	20	40	80	160
Number of iterations	7	40	269	1354
Execution time	0.39274	4.9762	117.139	2360.7
Error	0.04528	0.01746	0.00665	0.00357

Table 5.14: Error, number of iterations and execution time for the accelerated primal-dual algorithm for various spatial steps with optimal values $(\tau_0, r_0) = (h/2, h/4)$.

$N_x = N_y$	20	40	80	160
Number of iterations	2	6	21	65
Execution time	0.2434	1.2775	12.3118	145.835
Error	0.046120	0.02411	0.01062	0.00476

Table 5.15: Error, number of iterations and execution time for the regularization method for various spatial steps with optimal value $\varepsilon = \dfrac{800}{N_x^2} = 200h^2$.

$N_x = N_y$	20	40	80	160
Number of iterations	8	23	85	281
Execution time	0.5775	4.97562	70.8981	926.449
Error	0.06159	0.02216	0.00863	0.00409

Table 5.16: Error, number of iterations and execution time for the augmented Lagrangian algorithm for various spatial steps with constant penalty parameter $r = 10$.

5.2.5 Time dependent case

We now consider the Bingham problem in the time dependent case, i.e. (5.2.1) with αu replaced by $\partial_t u$. We build an exact solution as previously with $u(t,x,y) = \Phi(t,r)\begin{pmatrix}-y\\x\end{pmatrix}$ with $r = \sqrt{x^2+y^2}$. Then the problem can be written

$$\left(\partial_t\Phi(t,r) - \frac{\partial_r\big(\text{sgn }\partial_r\Phi(t,r)\big)}{r} - \frac{2\text{sgn }\partial_r\Phi(t,r)}{r^2}\right)\begin{pmatrix}-y\\x\end{pmatrix} = f. \tag{5.2.38}$$

An exact solution is built under the form

- For $0 \le r \le 1/6$,

$$\text{sgn }\partial_r\Phi = (12r - 36r^2)^2, \quad \Phi = t, \quad f = \big(1 - 2\times 12^2(1-3r)(2-9r)\big)\begin{pmatrix}-y\\x\end{pmatrix}.$$

- For $1/6 \le r \le 1/3$,

$$\text{sgn }\partial_r\Phi = 1, \quad \Phi = 6rt, \quad f = \big(6r - 2/r^2\big)\begin{pmatrix}-y\\x\end{pmatrix}.$$

- For $1/3 \le r \le 1/2$,

$$\text{sgn}\,\partial_r\Phi = \cos(\pi(6r-2)), \quad \Phi = 2t,$$
$$f = \left(2 + \frac{6\pi\sin(\pi(6r-2))}{r} - \frac{2\cos(\pi(6r-2))}{r^2}\right)\begin{pmatrix}-y\\x\end{pmatrix}.$$

- For $1/2 \le r \le 5/6$,

$$\text{sgn }\partial_r\Phi = -1, \quad \Phi = (5-6r)t, \quad f = \big(5 - 6r + 2/r^2\big)\begin{pmatrix}-y\\x\end{pmatrix}.$$

- For $5/6 \le r \le 1$,

$$\text{sgn }\partial_r\Phi = -\frac{1+\cos(\pi(6r-5))}{2}, \quad \Phi = 0,$$

$$f = \left(\frac{-3\pi\sin(\pi(6r-5))}{r} + \frac{1+\cos(\pi(6r-5))}{r^2}\right)\begin{pmatrix}-y\\x\end{pmatrix}.$$

- For $1 \le r$,

$$\text{sgn}\,\partial_r\Phi = 0, \quad \Phi = 0, \quad f = 0.$$

Comparison between various values of the timestep

Considering that the problem is of parabolic type we have a natural value of the timestep

$$\Delta t \sim h^2\frac{[Du]}{[\sigma]}. \tag{5.2.39}$$

N_t	20	40	80	160	320	640
N° of iterations	294	314	316	414	653	1162
Error estimate	0.2384	0.2815	0.3535	0.3586	0.3245	0.2294

Table 5.17: Time dependent model with $\tau = r = \frac{h}{3}$, for various values of the timestep

N_t	20	40	80	160	320	640
N° of iterations	57	55	84	160	320	640
Error	0.44292	0.4585	0.4681	0.4730	0.4755	0.4768

Table 5.18: Time dependent model with $\tau = \frac{h\Delta t}{3}$, $r = \frac{h}{3\Delta t}$, for various values of the timestep

We consider two choices $\tau = r = \frac{h}{3}$ or $\tau = \frac{h\Delta t}{3}$, $r = \frac{h}{3\Delta t}$. We take $t_{max} = 0.5$, $h = 1/20$, $\varepsilon_{tol} = 10^{-3}$. According to Tables 5.17 and 5.18 we can see that the first choice is better in terms of error, but the second one is less costly in terms of iterations. In both cases the value of the error depends very little on the timestep, this maybe due to the fact that the solution u is linear in time. More experiments would be necessary in order to better achieve the balance between τ and r, and decide whether the first or the second choice is the best.

5.3 Compressible Euler equations with Bingham viscoplasticity

5.3.1 Numerical scheme for the hyperbolic part

As described in the introduction, the compressible Euler equations with Bingham viscoplasticity can be solved by the splitting method. For the Euler hyperbolic part we have to solve

$$\partial_t \rho + \operatorname{div}(\rho u) = 0, \tag{5.3.1}$$

$$\partial_t(\rho u) + \operatorname{div}(\rho u \otimes u) + \nabla p = 0, \tag{5.3.2}$$

with $p = p(\rho)$. We shall take $p(\rho) = \rho^2/2$. This can be written as

$$\partial_t U + \operatorname{div}(F(U)) = 0, \tag{5.3.3}$$

or

$$\partial_t U + \partial_x(F_x(U)) + \partial_y(F_y(U)) = 0, \tag{5.3.4}$$

where $U = (\rho, \rho u)^T$ and $F(U) = (\rho u, \rho u \otimes u - pI)^T$. In order to solve the system (5.3.3) we use the finite volume method, which at first order consists in updating the value of U_i^n corresponding to each cell Q_i by the formula

$$U_i^{n+1} = U_i^n - \frac{\Delta t}{|Q_i|} \sum_{j \in N_i} |\Gamma_{ij}| F_{ij}, \tag{5.3.5}$$

where Δt is the timestep, $|Q_i|$ is the area of Q_i, $|\Gamma_{ij}|$ is the length of Γ_{ij}, N_i is the set of indices j corresponding to cells Q_j having a common interface Γ_{ij} with Q_i, and F_{ij} is a numerical flux between the cells Q_i, Q_j.

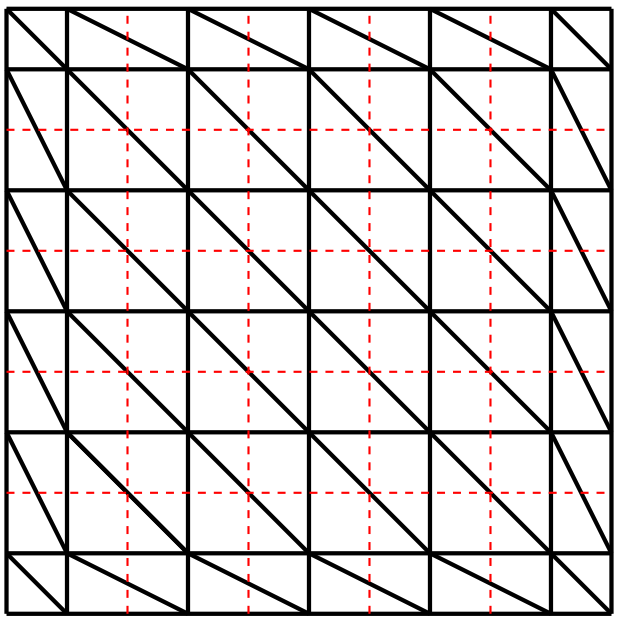


Figure 5.7: Primal mesh (black) and dual mesh(red) in the case without degrees of freedom on the boundary

Since (5.3.3) is conservative, it is natural to ask that (5.3.5) is conservative as well, i.e. we assume that

$$F_{ij} = -F_{ji}. \tag{5.3.6}$$

This ensures that $\sum_i |Q_i| U_i^n$ is time independent. We take the numerical flux of the form

$$F_{ij} = F(U_i, U_j, n_{ij}), \tag{5.3.7}$$

which means that given the direction n_{ij} we have to solve the system (5.3.3) in the direction n_{ij}. The consistency of the numerical scheme is that $F(U, U, n) = n_x F_x(U) + n_y F_y(U)$. More details can be found in [11].

Formulation with dual meshes

The location of the degrees of freedom are the (x_i, y_j) as defined by (5.2.3)-(5.2.7). The primal mesh is the FEM mesh made of triangles with (x_i, y_j) as nodes. The dual mesh is the FV mesh made of rectangles around each (x_i, y_j). The case with or without degrees of freedom on the boundary (depending on the type of boundary conditions) are shown on Figures 5.7 and 5.8.

5.3.2 Numerical results for the 2D Euler/Bingham model

We now consider the 2D Euler transport together with the viscoplastic model Bingham rheology

$$\partial_t \rho + \operatorname{div}(\rho u) = 0, \tag{5.3.8}$$

$$\partial_t (\rho u) + \operatorname{div}(\rho u \otimes u + p(\rho) I) - \operatorname{div}\left(\frac{Du}{|Du|}\right) = f, \tag{5.3.9}$$

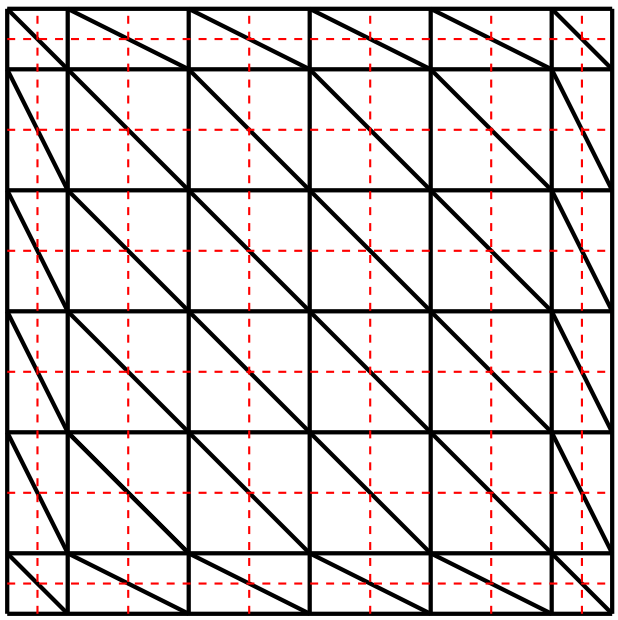


Figure 5.8: Primal mesh (black) and dual mesh(red) in the case of degrees of freedom on the boundary

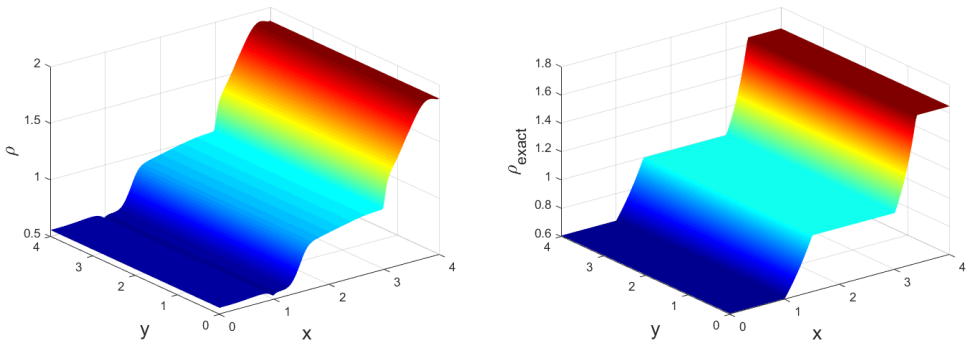


Figure 5.9: Euler/Bingham model: approximate and exact density ρ

with $p(\rho) = \frac{1}{2}\rho^2$.
We take the exact solution depending only on x

$$u = \begin{pmatrix} u_x(x) \\ 0 \end{pmatrix}, \tag{5.3.10}$$

where u_x, ρ and f_x are those of the one-dimensional solution given by (5.1.2), (5.1.6), (5.1.10). Slip boundary conditions are applied. The results are shown on Figures 5.9, 5.10, 5.11.
Number of time steps: 273.
Error: $\|(\rho, u_x, u_y)_{approx} - (\rho, u_x, u_y)_{ex}\|_{L^2} = 0.733$.

As we can observe on Table 5.19, by choosing the optimal value of ε_{tol} the scheme is first-order accurate with respect to the number of points in each direction, as with the 1d code (Table 5.20). This is true even though $u_{y_{exact}} = 0$ but $u_{y_{approx}}$ is not zero.

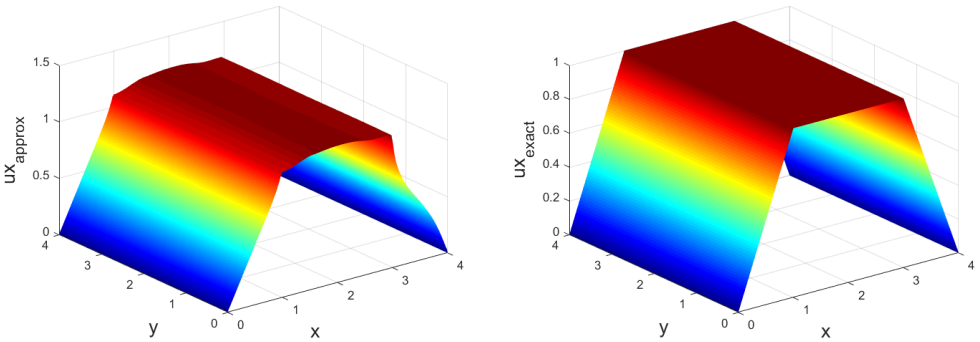


Figure 5.10: Euler/Bingham model: approximate and exact velocity u_x

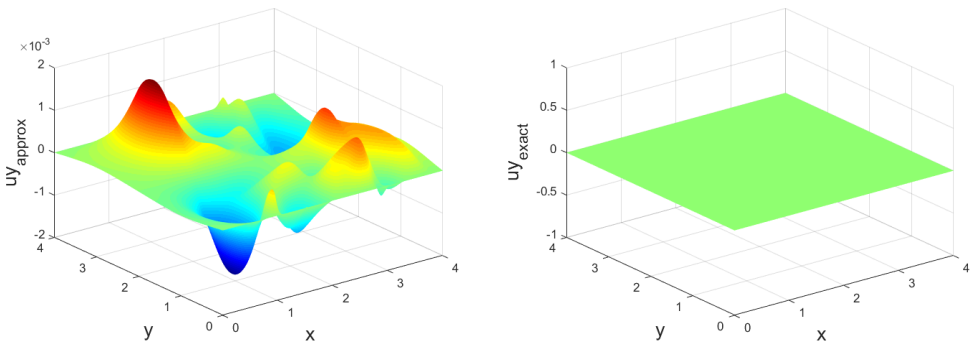


Figure 5.11: Euler/Bingham model: approximate and exact velocity u_y

N	error	number of iterations	ε_{tol}
20	0.763	31	10^{-2}
40	0.2219	60	10^{-4}
80	0.11	120	10^{-5}
160	$6.706e^{-2}$	241	10^{-6}
320	$3.817e^{-2}$	482	10^{-7}

Table 5.19: Error and number of iterations with the 2d code

N	error estimate
32	$7.07e^{-2}$
75	$4.541e^{-2}$
150	$2.695e^{-2}$
300	$3.87e^{-3}$
600	$3.03e^{-3}$
1200	$1.849e^{-3}$

Table 5.20: Error obtained with the 1d code

Chapter 6

A lubrication equation for a simplified model of shear-thinning fluid

This short chapter has been conducted independently from the other parts of this thesis. It is extracted from a manuscript which was submitted to ESAIM review: ProcS, a joint work with two other PhD students Khawla Msheik and Meissa M'Baye under supervision of François James. This proceeding is a part of a project launched in CEMRACS 2019.

6.1 Introduction

The lubrication equation is quite a classical simplification of the incompressible Navier-Stokes system. It is obtained for thin films of fluid, when viscous effects balance the pressure force. This occurs for instance for thin films of oil, hence the name of the equation. The study of this approximation goes back to Reynolds in 1886 [48]. Several scalings are involved to obtain this model. First the aspect ratio between the thickness of the film and the characteristic length of the substrate must be small, say δ. Simultaneously, the time scale has to be of order $1/\delta$. This is the so-called long wave regime, and is classically used in the shallow-water approximation. The lubrication equation requires another assumption of balance between the viscous effects and the pressure effects, which amounts to neglect all kinematic effects. This simplified flow is known as the Stokes flow. The lubrication equation itself is then obtained by integration over the fluid thickness.

We are interested here in the lubrication model for a class of non Newtonian fluids. Several fluids are known to depart from the usual Newtonian rheology, where the deviatoric stress tensor is a linear function of the strain rate tensor, thus defining the dynamical viscosity of the fluid. The lubrication equation for Newtonian fluids has been studied for instance by Huppert [35]. Non Newtonian fluids arise in several applications in engineering, biology, geophysics... In particular, viscoplastic or pseudoplastic fluids are involved in various geological problems, for instance lava flows, mudslides and avalanches. We refer to [1] for a review on the subject. A model which is widely used is the so-called Bingham-plastic model. This model involves a yield stress, namely a threshold on strain rate: for values of the strain rate above this threshold the fluid behaves like a viscous fluid, for values below, it looks like a solid. This can be thought of as an infinite viscosity fluid. We refer to the papers by Liu and Mei [41] and Balmforth et al. [2] for the study of such fluids in the lubrication approximation. Both papers contain also a complete bibliography. Liu and Mei also introduced in [40] a perturbed Bingham model, which is actually a two viscosities model, with a high viscosity for small deformations. When this viscosity goes to ∞ the Bingham model is

recovered, thus giving a fluid mechanics interpretation of this solid behaviour.

This is precisely the two viscosities model we investigate here. First we describe the mathematical model we use, namely the incompressible Navier-Sokes equations in a time-dependent domain, since we consider a free-boundary problem. In particular we explain in some details all the scalings involved. Next, we turn to the lubrication equation itself, which is a one-dimensional equation, obtained by averaging the previous ones along the thickness. Finally we provide a few numerical illustrations based on a finite volume scheme.

6.2 Mathematical model

In this section we set up the model. The starting point is the incompressible Navier-Stokes system. We limit ourselves in this paper to the two-dimensional case, thus aiming at a one-dimensional lubrication equation. Similar computations can be performed in three space dimensions. The domain we consider is Ω_t defined by $f_b(x) < z < \varphi(t,x)$, for $t > 0$ and $x \in (-\infty, +\infty)$, where f_b is given topography, and φ is a free surface. The notation we use is gathered in Figure 6.1.

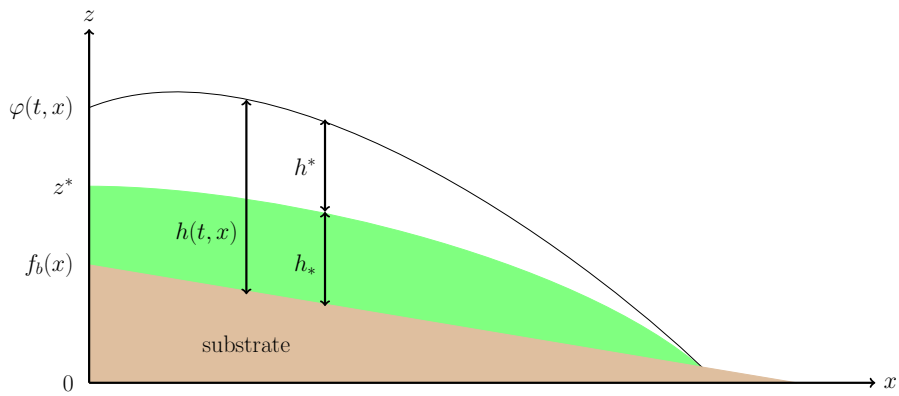


Figure 6.1: Notation for the two viscosities fluid: φ is the free surface; f_b is the topography of the substrate; z^* is the ordinate which separates "small deformations" (white zone) from "large deformations" (green zone), see Section 6.3 below. We introduce the thicknesses $h = \varphi - f_b$, $h^* = \varphi - z^*$, $h_* = h - h^*$.

The incompressible Navier-Stokes equations are

$$\partial_x u + \partial_z v = 0, \tag{6.2.1}$$

$$\partial_t u + u\partial_x u + v\partial_z u = -\frac{1}{\rho}\big(\partial_x p + \partial_x \tau_{xx} + \partial_z \tau_{xz}\big), \tag{6.2.2}$$

$$\partial_t v + u\partial_x v + v\partial_z v = -g - \frac{1}{\rho}\big(\partial_z p + \partial_x \tau_{zx} + \partial_z \tau_{zz}\big), \tag{6.2.3}$$

where ρ is the density of the fluid, $U = (u, v)$ is the velocity field, and the stress tensor σ is written as the sum of a volumetric stress tensor, involving the pressure p, and a deviatoric stress tensor τ:

$$\sigma = -p\mathrm{Id} + \tau, \qquad \tau = \begin{pmatrix} \tau_{xx} & \tau_{xz} \\ \tau_{zx} & \tau_{zz} \end{pmatrix}.$$

The density ρ is assumed to be constant here, and the tensor σ will be defined in Section 6.2.1 below.

Boundary conditions are:

$z = \varphi$: fluid-atmosphere interface. We have continuity of the stress tensor at the free surface, together with a kinematic boundary condition. Since the atmosphere can be viewed as an ideal fluid, the stress tensor can be taken equal to zero above φ. Hence we get

$$\sigma \cdot n|_\varphi = (-p\mathrm{Id} + \tau) \cdot n|_\varphi = 0, \qquad \partial_t \varphi + u_\varphi \partial_x \varphi = v_\varphi. \tag{6.2.4}$$

$z = f_b$: interface between the fluid and the substrate, which is fixed. This is a material interface, on which we have the no-slip boundary condition

$$u|_{f_b} = u_b, \quad v|_{f_b} = v_b. \tag{6.2.5}$$

Here (u_b, v_v) is the so-called basal velocity. Often in fluid mechanics the basal velocity is zero, but for geophysical applications it can actually be the driving force, and thus depend on (t, x).

6.2.1 Rheology

For a fluid, the deviatoric stress tensor τ is usually a function of the strain rate tensor

$$\dot{\varepsilon} = \begin{pmatrix} \dot{\varepsilon}_{xx} & \dot{\varepsilon}_{xz} \\ \dot{\varepsilon}_{zx} & \dot{\varepsilon}_{zz} \end{pmatrix} = \frac{1}{2}(\nabla U + \nabla U^T) = \frac{1}{2} \begin{pmatrix} 2\partial_x u & \partial_x v + \partial_z u \\ \partial_x v + \partial_z u & 2\partial_z v \end{pmatrix}. \tag{6.2.6}$$

A Newtonian fluid is characterized by a linear relation, defining the viscosity of the fluid, which is assumed here to be isotropic and constant. Therefore we introduce the dynamical viscosity coefficient μ, and define the Newtonian stress tensor by

$$\tau_N = 2\mu\dot{\varepsilon} = 2\rho\nu\dot{\varepsilon},$$

where $\nu = \mu/\rho$ is the kinematic viscosity.

Fluids that do not follow this kind of constitutive law are non-Newtonian. In the general case, the material invariance principle implies that the stress tensor depends only on the similarity invariants of the strain rate tensor, in particular the coefficients of its characteristic polynomial. In dimension 2 there are only two such coefficients $\dot{\varepsilon}_I$ and $\dot{\varepsilon}_{II}$. Namely $\dot{\varepsilon}_I$ is the trace of the matrix and $\dot{\varepsilon}_{II}$ its determinant. For an incompressible fluid, the trace is zero, and moreover we have

$$\dot{\varepsilon}_{II} = \dot{\varepsilon}_{xx}\dot{\varepsilon}_{zz} - \dot{\varepsilon}_{zx}\dot{\varepsilon}_{xz} = \partial_x u \partial_z v - \frac{1}{4}(\partial_x v + \partial_z u)^2 = -\big((\partial_x u)^2 + \frac{1}{4}(\partial_x v + \partial_z u)^2\big).$$

This allows to define the strain rate $\dot{\gamma}$ as

$$\dot{\gamma} = 2\sqrt{-\dot{\varepsilon}_{II}} = 2\sqrt{(\partial_x u)^2 + \frac{1}{4}(\partial_x v + \partial_z u)^2}. \tag{6.2.7}$$

In a similar way we can check that the Frobenius norm of $\dot{\varepsilon}$, that is $\|\dot{\varepsilon}\|^2 = \sum_{i,j}(\varepsilon_{ij})^2$, satisfies

$$\|\dot{\varepsilon}\|^2 = \dot{\gamma}^2/2. \tag{6.2.8}$$

A very sketchy illustration of the possible behaviours of non-Newtonian fluids is given in Figure 6.2. We will be mostly interested in this work in the so-called pseudoplastic case, that is the red curve in Figure 6.2, for which experimental evidence can be given, see [10]. This kind of models are also used in geophysics, see [6, 56, 60] We wish to give a simplified model for this pseudo-plastic

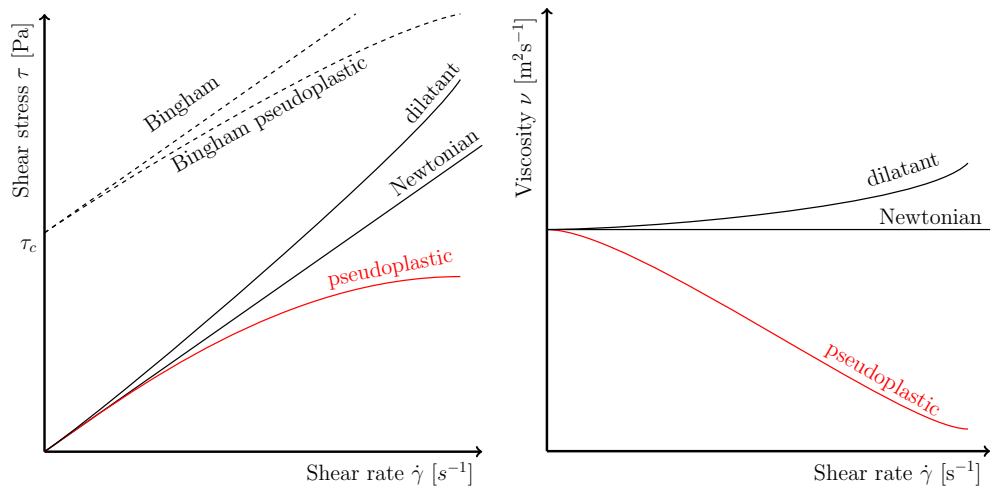


Figure 6.2: Qualitative behaviour of various types of fluids. Left: stress vs shear stress – Right: apparent viscosity vs shear stress. The Bingham type fluids can be viewed as enjoying infinite apparent viscosity below the threshold τ_c.

fluid, that allows to handle explicit computations. The main feature of this kind of fluids is a nonlinear viscosity, decreasing with the strain rate. Mimicking the Bingham model, which is based on a threshold on the shear stress, we consider a model with a threshold on the strain rate: the viscosity is equal to some large μ_B for small deformations, that is $\dot{\gamma} < \gamma_c$, where $\gamma_c > 0$ is a given constant, and to another value μ for large deformations, $\dot{\gamma} > \gamma_c$. Such models were introduced by Liu and Mei [40], and the limit case $\nu_B \to \infty$, which leads to a Bingham fluid, is studied in [41] and [2]. Notice that using (6.2.8) the threshold γ_c on $\dot{\gamma}$ can be replaced by a threshold $\gamma_c' = \gamma_c/\sqrt{2}$ on $\|\dot{\varepsilon}\|$.

A multidimensional formulation for these simplified pseudo-plastic fluids is therefore

$$\tau_{PP} = \begin{cases} 2\rho\nu_B\dot{\varepsilon} & \text{if } \|\dot{\varepsilon}\| \leqslant \gamma_c', \\ 2\rho\nu\dot{\varepsilon} + 2\rho(\nu_B - \nu)\gamma_c'\dfrac{\dot{\varepsilon}}{\|\dot{\varepsilon}\|} & \text{if } \|\dot{\varepsilon}\| > \gamma_c', \end{cases} \tag{6.2.9}$$

where we have introduced the kinematic viscosities ν and ν_B.

A particular limit case is $\nu_B \to \infty$, which leads to a Bingham type fluid. To view this, it is convenient to define the following quantities (see Figure 6.3 below for an illustration in 1 dimension)

$$\tau_c = \nu_B\gamma_c', \qquad \tau_* = (\nu_B - \nu)\gamma_c' = (1 - \nu/\nu_B)\tau_c, \tag{6.2.10}$$

so that definition (6.2.9) can be rewritten

$$\tau_{PP} = \begin{cases} 2\rho\nu_B\dot{\varepsilon} & \text{if } \|\dot{\varepsilon}\| \leqslant \tau_c/\nu_B, \\ 2\rho\nu\dot{\varepsilon} + 2\rho(1 - \nu/\nu_B)\tau_c\dfrac{\dot{\varepsilon}}{\|\dot{\varepsilon}\|} & \text{if } \|\dot{\varepsilon}\| > \tau_c/\nu_B. \end{cases} \tag{6.2.11}$$

It is clear on this formulation that the relevant limit is $\nu_B \to +\infty$, together with $\gamma_c' \to 0$, keeping $\nu_B\gamma_c' = \tau_c$. In doing so, we recover the classical Bingham stress tensor, with threshold τ_c:

$$\tau_{\text{Bing}} = \begin{cases} \text{any } \tau \text{ s.t. } \|\tau\| \leqslant \tau_c & \text{if } \dot{\varepsilon} = 0, \\ 2\rho\nu\dot{\varepsilon} + 2\rho\tau_c\dfrac{\dot{\varepsilon}}{\|\dot{\varepsilon}\|} & \text{if } \|\dot{\varepsilon}\| > 0. \end{cases}$$

Finally, notice that in the pseudo-plastic (or shear thinning) context, we consider $0 < \nu < \nu_B$, but similar computations can be performed in any case.

It is convenient for the scalings below to rewrite expression (6.2.11) using an equivalent kinematic viscosity ν_{eq}, which satisfies $\nu \leqslant \nu_{eq} \leqslant \nu_B$:

$$\tau_{PP} = \rho \nu_{eq} \dot{\varepsilon}, \qquad \text{where } \nu_{eq} = \begin{cases} 2\nu_B & \text{if } \|\dot{\varepsilon}\| \leqslant \gamma_c', \\ 2\nu + 2(1-\nu/\nu_B)\dfrac{\tau_c}{\|\dot{\varepsilon}\|} & \text{if } \|\dot{\varepsilon}\| > \gamma_c'. \end{cases} \tag{6.2.12}$$

6.2.2 Scalings

We introduce now the scaling laws, namely thin layer, or more precisely long wave approximation, and slow motion, in order to finally obtain the lubrication model. This kind of scalings is already present e.g. in [2] in the context of a visco-plastic fluid. Hence we propose the following family of scalings: we introduce a first set of characteristic scales, namely dimensions ℓ_0 and h_0, characteristic velocities u_0 and v_0, and a characteristic time t_0. The quantities ℓ_0 and u_0 correspond to the horizontal direction, h_0 and v_0 to the vertical one. The aspect ratio $\delta = h_0/\ell_0$ will be an important parameter, assumed to be small in the thin layer case. Dimensionless variables are then defined by

$$\begin{aligned} x &= \ell_0 \bar{x}, \quad z = h_0 \bar{z}, \quad t = t_0 \bar{t} \\ u &= u_0 \bar{u}, \quad v = v_0 \bar{v}. \end{aligned}$$

First, we rewrite the incompressibility equation (6.2.1) in the rescaled variables. We obtain

$$\frac{u_0}{\ell_0}\partial_{\bar{x}}\bar{u} + \frac{v_0}{h_0}\partial_{\bar{z}}\bar{v} = 0,$$

and following the least degeneracy principle [59], this implies $u_0/\ell_0 = v_0/h_0$, or equivalently $\ell_0/h_0 = u_0/v_0$. Thus $v_0/u_0 = \delta$, so that in the thin layer approximation v_0 is also small compared to u_0.

We turn now to the kinematic part of the equation. Using $u_0/\ell_0 = v_0/h_0$, we readily obtain

$$\partial_t u + u\partial_x u + v\partial_z u = \frac{u_0}{t_0}\partial_{\bar{t}}\bar{u} + \frac{u_0 v_0}{h_0}\bar{u}\partial_{\bar{x}}\bar{u} + \frac{u_0 v_0}{h_0}\bar{v}\partial_{\bar{z}}\bar{u}.$$

Once again we apply the least degeneracy principle and obtain $t_0 = \ell_0/u_0 = h_0/v_0$, or, as expected, $u_0 = \ell_0/t_0$ and $v_0 = h_0/t_0$. We proceed in the same way for the momentum equation in v and finally obtain

$$\partial_t u + u\partial_x u + v\partial_z u = \frac{u_0 v_0}{h_0}\left(\partial_{\bar{t}}\bar{u} + \bar{u}\partial_{\bar{x}}\bar{u} + \bar{v}\partial_{\bar{z}}\bar{u}\right) = \delta\frac{u_0^2}{h_0}\left(\partial_{\bar{t}}\bar{u} + \bar{u}\partial_{\bar{x}}\bar{u} + \bar{v}\partial_{\bar{z}}\bar{u}\right), \tag{6.2.13}$$

$$\partial_t v + u\partial_x v + v\partial_z v = \frac{u_0 v_0}{\ell_0}\left(\partial_{\bar{t}}\bar{v} + \bar{u}\partial_{\bar{x}}\bar{v} + \bar{v}\partial_{\bar{z}}\bar{v}\right) = \delta^2\frac{u_0^2}{h_0}\left(\partial_{\bar{t}}\bar{v} + \bar{u}\partial_{\bar{x}}\bar{v} + \bar{v}\partial_{\bar{z}}\bar{v}\right), \tag{6.2.14}$$

where we have emphasized the aspect factor $\delta = h_0/\ell_0 = v_0/u_0$.

Following Balmforth [2], we rescale the pressure and the stress tensor by

$$p = \rho g h_0 \bar{p}, \qquad \tau = \rho\nu\frac{u_0}{h_0}\bar{\tau}. \tag{6.2.15}$$

We can write now the rescaled version of the Navier-Stokes momentum equations (6.2.2) and (6.2.3):

$$\delta\frac{u_0^2}{h_0}\left(\partial_{\bar{t}}\bar{u} + \bar{u}\partial_{\bar{x}}\bar{u} + \bar{v}\partial_{\bar{z}}\bar{u}\right) = -\delta g\partial_{\bar{x}}\bar{p} + \nu\frac{u_0}{h_0^2}\left(\delta\partial_{\bar{x}}\bar{\tau}_{xx} + \partial_{\bar{z}}\bar{\tau}_{xz}\right), \tag{6.2.16}$$

$$\delta^2\frac{u_0^2}{h_0}\left(\partial_{\bar{t}}\bar{v} + \bar{u}\partial_{\bar{x}}\bar{v} + \bar{v}\partial_{\bar{z}}\bar{v}\right) = -g\partial_{\bar{z}}\bar{p} - g + \nu\frac{u_0}{h_0^2}\left(\delta\partial_{\bar{x}}\bar{\tau}_{xz} + \partial_{\bar{z}}\bar{\tau}_{zz}\right). \tag{6.2.17}$$

At this stage, we introduce two classical dimensionless quantities, namely the Froude and Reynolds numbers, defined from the characteristic *horizontal* velocity u_0, the *vertical* extension h_0, and the viscosity for *large deformations* ν:

$$\frac{1}{Fr^2} = \frac{gh_0}{u_0^2}, \qquad \frac{1}{Re} = \frac{\nu}{u_0 h_0}. \tag{6.2.18}$$

We divide the previous two equations by u_0^2/h_0, and noticing that $\tau = \rho Re u_0^2 \bar{\tau}$, we obtain

$$\delta\left(\partial_{\bar{t}}\bar{u} + \bar{u}\partial_{\bar{x}}\bar{u} + \bar{v}\partial_{\bar{z}}\bar{u}\right) = -\frac{\delta}{Fr^2}\partial_{\bar{x}}\bar{p} + \frac{1}{\delta Re}\left(\delta\partial_{\bar{x}}\bar{\tau}_{xx} + \partial_{\bar{z}}\bar{\tau}_{xz}\right), \tag{6.2.19}$$

$$\delta^2\left(\partial_{\bar{t}}\bar{v} + \bar{u}\partial_{\bar{x}}\bar{v} + \bar{v}\partial_{\bar{z}}\bar{v}\right) = -\frac{1}{Fr^2}\partial_{\bar{z}}\bar{p} - \frac{1}{Fr^2} + \frac{1}{Re}\left(\delta\partial_{\bar{x}}\bar{\tau}_{xz} + \partial_{\bar{z}}\bar{\tau}_{zz}\right). \tag{6.2.20}$$

The idea now is to send δ to zero, thus implementing the thin layer assumption, but in a regime where the Reynolds number Re is kept of order 1, together with a balance between viscosity and gravity forces. Therefore we set

$$Fr^2 = \delta Re. \tag{6.2.21}$$

This readily gives

$$u_0 = \frac{gh_0^3}{\ell_0\nu} = \delta\frac{gh_0^2}{\nu}, \tag{6.2.22}$$

which is the scaling proposed in [2]. It introduces another characteristic velocity, namely $u_0' = (gh_0^2)/\nu$. The latter equality shows that this is indeed a slow motion scaling, thus we meet the initial requirement.

Inserting (6.2.21) in equations (6.2.19) and (6.2.20), and keeping only the dominant terms of order δ^{-1} gives first the dimensionless Stokes equation

$$-\partial_{\bar{x}}\bar{p} = \partial_{\bar{z}}\bar{\tau}_{xz}, \tag{6.2.23}$$

then the dimensionless hydrostatic relation for the pressure

$$\partial_{\bar{z}}\bar{p} = -1. \tag{6.2.24}$$

Now we compute $\bar{\tau}$ from (6.2.12). We start by rewriting $\dot{\varepsilon}$ in rescaled variables

$$\dot{\varepsilon} = \frac{1}{2}\begin{pmatrix} 2\frac{u_0}{\ell_0}\partial_{\bar{x}}\bar{u} & \frac{v_0}{\ell_0}\partial_{\bar{x}}\bar{v} + \frac{u_0}{h_0}\partial_{\bar{z}}\bar{u} \\ \frac{v_0}{\ell_0}\partial_{\bar{x}}\bar{v} + \frac{u_0}{h_0}\partial_{\bar{z}}\bar{u} & 2\frac{v_0}{h_0}\partial_{\bar{z}}\bar{v} \end{pmatrix} = \frac{1}{2}\frac{u_0}{h_0}\begin{pmatrix} 2\delta\partial_{\bar{x}}\bar{u} & \delta^2\partial_{\bar{x}}\bar{v} + \partial_{\bar{z}}\bar{u} \\ \delta^2\partial_{\bar{x}}\bar{v} + \partial_{\bar{z}}\bar{u} & 2\delta\partial_{\bar{z}}\bar{v} \end{pmatrix}. \tag{6.2.25}$$

From this we easily deduce

$$\bar{\tau} = \frac{1}{2}\frac{\nu_{eq}}{\nu}\begin{pmatrix} 2\delta\partial_{\bar{x}}\bar{u} & \delta^2\partial_{\bar{x}}\bar{v} + \partial_{\bar{z}}\bar{u} \\ \delta^2\partial_{\bar{x}}\bar{v} + \partial_{\bar{z}}\bar{u} & 2\delta\partial_{\bar{z}}\bar{v} \end{pmatrix} \underset{\delta\to 0}{\longrightarrow} \frac{1}{2}\frac{\nu_{eq}}{\nu}\begin{pmatrix} 0 & \partial_{\bar{z}}\bar{u} \\ 0 & 0 \end{pmatrix}. \tag{6.2.26}$$

We define a dimensionless equivalent viscosity by $\bar{\nu}_{eq} = \nu_{eq}/\nu$, and rewrite equation (6.2.23)

$$\partial_{\bar{z}}\left(\bar{\nu}_{eq}\partial_z\bar{u}\right) = -\partial_{\bar{x}}\bar{p}. \tag{6.2.27}$$

We turn now to the expression of $\bar{\nu}_{eq}$. We first notice that, using (6.2.25)

$$\|\dot{\varepsilon}\| = \frac{u_0}{h_0}\sqrt{2}\sqrt{\delta^2(\partial_{\bar{x}}\bar{u})^2 + \frac{1}{4}(\delta^2\partial_{\bar{x}}\bar{v} + \partial_{\bar{z}}\bar{u})^2} \underset{\delta\to 0}{\longrightarrow} \frac{1}{\sqrt{2}}\frac{u_0}{h_0}|\partial_{\bar{z}}\bar{u}|. \tag{6.2.28}$$

Hence the condition $\|\dot{\varepsilon}\| > \gamma_c'$ leads us to define a dimensionless threshold $\bar{\gamma}_c = (\sqrt{2}h_0/u_0)\gamma_c' = (h_0/u_0)\gamma_c$, so that the condition $\|\dot{\varepsilon}\| > \gamma_c'$ becomes $|\partial_{\bar{z}}\bar{u}| > \bar{\gamma}_c$. Thus we get

$$\bar{\nu}_{eq} = \frac{\nu_{eq}}{\nu} = \begin{cases} 2\dfrac{\nu_B}{\nu} & \text{if } |\partial_{\bar{z}}\bar{u}| \leqslant \bar{\gamma}_c, \\ 2 + 2(1-\nu/\nu_B)\dfrac{\tau_c}{\nu}\dfrac{h_0}{u_0}\dfrac{\sqrt{2}}{|\partial_{\bar{z}}\bar{u}|} & \text{if } |\partial_{\bar{z}}\bar{u}| > \bar{\gamma}_c. \end{cases} \tag{6.2.29}$$

We introduce a dimensionless viscosity $\bar{\nu}_B$ and a dimensionless yield stress B by setting

$$\bar{\nu}_B = \frac{\nu_B}{\nu} \geqslant 1, \qquad B = \frac{\sqrt{2}\tau_c h_0}{\nu u_0}, \tag{6.2.30}$$

so that the dimensionless deviatoric stress tensor becomes (see Figure 6.3)

$$\bar{\tau}_{xz} = \begin{cases} \bar{\nu}_B \partial_{\bar{z}}\bar{u} & \text{if } |\partial_{\bar{z}}\bar{u}| \leqslant \bar{\gamma}_c, \\ \partial_{\bar{z}}\bar{u} + (1-1/\bar{\nu}_B)B\,\mathrm{sgn}(\partial_{\bar{z}}\bar{u}) & \text{if } |\partial_{\bar{z}}\bar{u}| > \bar{\gamma}_c. \end{cases} \tag{6.2.31}$$

This is the model proposed by Liu and Mei in [40].

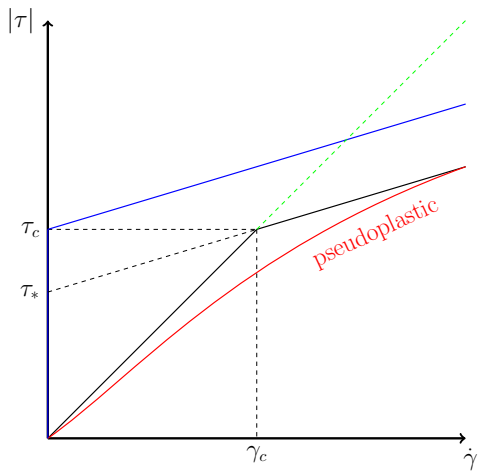


Figure 6.3: Simplified shear-thinning model. We consider a piecewise linear approximation (in black) of the "theoretical" pseudoplastic law (in red). Parameters τ_* and τ_c are defined by (6.2.10). The blue curve is the Bingham limit: $\nu_B \to +\infty$, $\gamma_c \to 0$ with $\gamma_c\nu_B = \tau_c$. The green dashed line is the pure Newtonian limit $\nu_B \to \nu$.

As concerns the boundary conditions, we notice that the no-slip and kinematic boundary conditions remain unchanged by the scaling. In contrast, the continuity of the stress tensor across the free surface φ is greatly simplified. Recalling that $\varphi = h_0\bar{\varphi}$, we indeed obtain

$$\sigma \cdot n = \begin{pmatrix} -p\partial_x\eta + \tau_{xx}\partial_x\varphi - \tau_{xz} \\ -p + \tau_{xz}\partial_x\varphi - \tau_{zz} \end{pmatrix} = \rho \begin{pmatrix} -\delta\bar{p}gh_0\partial_{\bar{x}}\bar{\varphi} + \delta\nu\dfrac{u_0}{h_0}\bar{\tau}_{xx}\partial_{\bar{x}}\bar{\varphi} - \nu\dfrac{u_0}{h_0}\bar{\tau}_{xz} \\ -\bar{p}gh_0 + \delta\nu\dfrac{u_0}{h_0}\bar{\tau}_{xz}\partial_{\bar{x}}\bar{\varphi} - \nu\dfrac{u_0}{h_0}\bar{\tau}_{zz} \end{pmatrix}.$$

Now making use of (6.2.26), we obtain

$$\sigma \cdot n = \rho \begin{pmatrix} \delta\big((\delta\nu_{eq}\dfrac{u_0}{h_0}\partial_{\bar{x}}\bar{u} - gh_0\bar{p})\partial_{\bar{x}}\bar{\varphi} - \delta\nu_{eq}\dfrac{u_0}{2h_0}\partial_{\bar{z}}\bar{v}\big) - \nu_{eq}\dfrac{u_0}{2h_0}\partial_{\bar{z}}\bar{u} \\ \\ \delta\nu_{eq}\dfrac{u_0}{2h_0}\big(\delta^2\partial_{\bar{x}}\bar{v} + (\partial_{\bar{z}}\bar{u})\partial_{\bar{x}}\bar{\varphi} - \partial_{\bar{z}}\bar{v}\big) - gh_0\bar{p} \end{pmatrix}.$$

Letting δ go to zero gives therefore

$$[\bar{p}]|_{\bar{\varphi}} = 0, \qquad [\partial_{\bar{z}}\bar{u}]|_{\bar{\varphi}} = 0, \tag{6.2.32}$$

where $[g]_{\bar{\varphi}}$ is the jump of some function g across $\bar{\varphi}$. In other words, we recover separately the continuity of the pressure and the continuity of $\partial_z u$.

6.3 Lubrication equation

The so-called lubrication equation is obtained by integrating equations (6.2.1) along the vertical direction. The long wave and slow motion assumptions imply that we obtain a single nonlinear equation on the depth φ. Similar computations were performed by Liu and Mei [40], for a two-viscosity model, in order to justify the Bingham case, which corresponds to $\nu_B \to \infty$ in our context. For Bingham fluids, we refer to Liu and Mei [41], and more recently to Balmforth [2]. The final equation is obtained through three steps we present in detail now.

We recall the equations we obtained in the preceding section, dropping the bars for clarity. First we have the hydrostatic relation

$$\partial_z p = -1, \quad f_b \leqslant z \leqslant \varphi. \tag{6.3.1}$$

Next, the dimensionless Stokes equation (6.2.23)

$$\partial_z(\tau_{xz}) = -\partial_x p, \quad f_b \leqslant z \leqslant \varphi, \tag{6.3.2}$$

where τ_{xz} is the dimensionless deviatoric stress tensor defined by (6.2.31).

These equations are coupled with the following boundary conditions (in these relations, t and x are hidden parameters):

- on the free surface $z = \varphi$
$$p(\varphi) = 0, \quad \partial_z u(\varphi) = 0, \tag{6.3.3}$$
- on $z = f_b$
$$u(f_b) = u_b, \quad v(f_b) = v_b. \tag{6.3.4}$$

Concerning first the pressure, using the boundary condition on the free surface we obtain the usual hydrostatic approximation

$$p(z) = \varphi - z, \qquad f_b \leqslant z \leqslant \varphi. \tag{6.3.5}$$

The averaged equation we look for is obtained by integrating in z the incompressibility equation, or mass conservation,

$$\partial_x u + \partial_z v = 0.$$

This is quite classical, see e.g. [2] in the same slow motion context, or [36] for shallow water approximation. We obtain

$$v(t,x,\varphi) = v(t,x,f_b) - \int_{f_b}^{\varphi} \partial_x u(z)\, dz \tag{6.3.6}$$

$$= v(t,x,f_b) - \partial_x \left(\int_{f_b}^{\varphi} u(z)\, dz \right) + u(t,x,\varphi)\partial_x\varphi - u(t,x,f_b)\partial_x f_b. \tag{6.3.7}$$

The kinematic boundary condition on $z = \varphi$ leads to $v(t,x,\varphi) - u(t,x,\varphi)\partial_x\varphi = \partial_t\varphi = \partial_t h$ in the first equation. For $z = f_b$, we make use of the no-slip boundary condition (6.3.4), to obtain the following averaged equation

$$\partial_t h + \partial_x \left(\int_{f_b}^{\varphi} u(z)\, dz \right) = v_b. \tag{6.3.8}$$

The flux $\int_{f_b}^{\varphi} u(z)\, dz$ can be computed explicitly as a function of φ, by integrating twice equation (6.3.2).

The first step towards the computation of the flux is to obtain the vertical velocity profile. The general structure of this profile is as follows. We have $\tau_{xz} = F(\partial_z u)$, where F is a continuous, one-to-one, increasing function, with $F(0) = 0$, see (6.2.31) and Figure 6.3. From (6.3.2) and (6.3.5) we are led to solve $\partial_z\big(F(\partial_z u)\big) = \partial_x\varphi$. Since $F(\partial_z u) = 0$ for $z = \varphi$ (or equivalently $\partial_z u = 0$) we get $F(\partial_z u) = \partial_x\varphi(z - \varphi)$, so that $F(\partial_z u)$ is monotone (increasing if $\partial_x\varphi \geqslant 0$, decreasing if not). Because F is increasing, $\partial_z u$ is monotone as well, in particular, since $\partial_z u = 0$ for $z = \varphi$, its sign remains constant. Therefore $|\partial_z u|$ is decreasing in z (increasing with depth).

The threshold in formula (6.2.31) eventually splits the fluid in two layers. Let z^* be defined by $|\partial_z u(z^*)| = \gamma_c$. Provided $z^* \in]f_b, \varphi[$ (see below for precise formulas), we have a "small deformation", that is $|\partial_z u(z)| < \gamma_c$, region for $z \in]z^*, \varphi[$, because $|\partial_z u|$ is decreasing from 0 for increasing depth. Similarly for $z \in]f_b, z^*[$ we have $|\partial_z u(z)| > \gamma_c$, so that finally, according to (6.2.31), the velocity is ruled by the system of equations

$$\begin{aligned} \nu_B \partial_{zz} u &= \partial_x\varphi, \qquad && z^* \leqslant z \leqslant \varphi, \\ \partial_{zz} u &= \partial_x\varphi, \qquad && f_b \leqslant z \leqslant z^*, \end{aligned}$$

where for the second equation we have used that $\partial_z u$ has a constant sign. These equations are complemented with the boundary conditions

$$\partial_z u = 0, \quad z = \varphi\,; \qquad u = 0, \quad z = f_b.$$

Notice that the curve $z = z^*$ is not a physical interface, yet we have continuity of the stress tensor, or equivalently here continuity of $\partial_z u$.

Now the computations are quite easy. We integrate once the first equation between φ and z^*, to obtain

$$\partial_z u = \frac{1}{\nu_B}\partial_x\varphi(z - \varphi).$$

This leads to

$$\dot\gamma = |\partial_z u_S| = \frac{1}{\nu_B}|\partial_x\varphi|(\varphi - z),$$

so that the value of z^* and the thickness h^* of this layer are given by

$$z^* = \max\left(\varphi - \frac{B}{|\partial_x\varphi|}, f_b\right), \qquad h^* = \varphi - z^* = \min\left(\frac{B}{|\partial_x\varphi|}, h\right). \tag{6.3.9}$$

These definitions ensure that $z^* \geqslant f_b$ and $h^* \leqslant h$, and are valid for $\partial_x\varphi = 0$ with the convention $B/0 = \infty$. Notice that z^* can be equal to f_b for weak slopes (small $\partial_x\varphi$), or small depths (small h). Conversely, $z^* \to \varphi$ when $|\partial_x\varphi|$ goes to ∞.

Integrating once again between z^* and φ, we obtain the velocity profile for $\varphi \geqslant z \geqslant z^*$:

$$u(z) = \frac{1}{2\nu_B}\partial_x\varphi(\varphi - z)^2 + K,$$

where the constant K will be determined later. Notice for further use that by construction

$$\partial_z u(z^*) = \frac{1}{\nu_B}\partial_x\varphi(z^* - \varphi) = \gamma_c. \tag{6.3.10}$$

We turn now to the lower layer, $z^* \geqslant z \geqslant f_b$. The fluid here has dimensionless viscosity 1, and we use the boundary conditions (6.3.10) for $z = z^*$, and no slip (6.3.4) at $z = f_b$. First we get, using (6.3.10),

$$\partial_z u = \partial_x\varphi(z - z^*) - \frac{1}{\nu_B}\partial_x\varphi h^*,$$

next, integrating once again between f_b and z^*,

$$u = \frac{1}{2}\partial_x\varphi(z^* - z)^2 - \frac{1}{\nu_B}\partial_x\varphi h^* z + L,$$

where L is computed using (6.3.4), leading to

$$L = u_b - \frac{1}{2}\partial_x\varphi(z^* - f_b)^2 + \frac{1}{\nu_B}\partial_x\varphi h^* f_b,$$

so that

$$u = \frac{1}{2}\partial_x\varphi\left((z^* - z)^2 - (z^* - f_b)^2\right) - \frac{1}{\nu_B}\partial_x\varphi h^*(z - f_b) + u_b, \tag{6.3.11}$$

Finally, we use the continuity of the velocity at $z = z^*$ to obtain the constant K:

$$\frac{1}{2\nu_B}\partial_x\varphi(\varphi - z^*)^2 + K = -\frac{1}{2}\partial_x\varphi(z^* - f_b)^2 - \frac{1}{\nu_B}\partial_x\varphi h^*(z^* - f_b) + u_b.$$

The velocity profile is therefore given by

$$u(z) = \begin{cases} \dfrac{\partial_x\varphi}{2}\left((z^* - z)^2 - (z^* - f_b)^2\right) - \dfrac{\partial_x\varphi}{\nu_B}h^*(z - f_b) + u_b, & f_b \leqslant z \leqslant z^*, \\[2ex] \dfrac{\partial_x\varphi}{2}\left((\varphi - z)^2 - (\varphi - z^*)^2\right) - \dfrac{\partial_x\varphi}{2}(z^* - f_b)^2 - \dfrac{\partial_x\varphi}{\nu_B}h^*(z^* - f_b) + u_b, & z^* \leqslant z \leqslant \varphi. \end{cases} \tag{6.3.12}$$

Notice that for $\nu_B = 1$, easy computations show that the profile is the same in the two layers, namely $u = \dfrac{\partial_x\varphi}{2}(z - f_b)(z - f_b - 2h) + u_b$, which is as expected the usual parabolic profile for a Newtonian fluid.

On the other hand, letting $\nu_B \to +\infty$, and $\gamma_c \to 0$, keeping $\nu_B\gamma_c = \tau_c$, we recover formally the Bingham fluid velocity, as in Balmforth [2]:

$$u(z) = \begin{cases} -\dfrac{\partial_x\varphi}{2}\left((y^*)^2 - (y^* - (z - f_b))^2\right) + u_b, & f_b \leqslant z \leqslant z^*, \\[2ex] -\dfrac{\partial_x\varphi}{2}(y^*)^2 + u_b, & z^* \leqslant z \leqslant \varphi, \end{cases}$$

where we have set $y^* = z^* - f_b = h - h^*$.

It is now straightforward to obtain the flux in (6.3.8), since

$$\int_{f_b}^{\varphi} u(z)\,dz = \int_{f_b}^{z^*} u(z)\,dz + \int_{z^*}^{\varphi} u(z)\,dz.$$

We have on the one hand

$$\int_{f_b}^{z^*} u(z)\,dz = -\frac{\partial_x\varphi}{3}(y^*)^3 - \frac{\partial_x\varphi}{2\nu_B}h^*(y^*)^2 + u_b y^*,$$

on the other hand

$$\int_{z^*}^{\varphi} u_T(z)\,dz = -\frac{\partial_x\varphi}{3\nu_B}(h^*)^3 - \frac{\partial_x\varphi}{2}(y^*)^2 h^* - \frac{\partial_x\varphi}{\nu_B}y^*(h^*)^2 + u_b h^*.$$

Therefore the flux we are looking for is given by

$$\int_{f_b}^{\varphi} u(z)\,dz = -\frac{\partial_x\varphi}{3}\left((y^*)^3 + \frac{3}{2}\big(\frac{1}{\nu_B}+1\big)(y^*)^2 h^* + \frac{3}{\nu_B}y^*(h^*)^2 + \frac{1}{\nu_B}(h^*)^3\right) + u_b h. \tag{6.3.13}$$

It is easy once again to check on this formula that we recover the usual cubic flux $-\frac{\partial_x\varphi}{3}h^3$ for the Newtonian fluid $\nu_B = 1$. On the other hand the limit case $\nu_B \to \infty$ gives back Balmforth's formula

$$\int_{f_b}^{\varphi} u(z)\,dz = -\frac{\partial_x\varphi}{6}(y^*)^2(y^* - 3h).$$

Inserting (6.3.13) in the conservation equation (6.3.8) leads to the following advection-diffusion equation:

$$\partial_t h + \partial_x(u_b h) = v_b + \partial_x\left(D(h,\partial_x h)\partial_x(h+f_b)\right), \tag{6.3.14}$$

where

$$D(h,\partial_x h) = \frac{1}{3}\left((y^*)^3 + \frac{3}{2}\big(\frac{1}{\nu_B}+1\big)(y^*)^2 h^* + \frac{3}{\nu_B}y^*(h^*)^2 + \frac{1}{\nu_B}(h^*)^3\right) \tag{6.3.15}$$

and we recall the definitions of h^* from (6.3.9), and y^*

$$h^* = \min\left(\frac{B}{|\partial_x\varphi|}, h\right), \qquad y^* = h - h^* = z^* - f_b. \tag{6.3.16}$$

Notice that $0 < D(h,\partial_x h) \leqslant h^3/(3\nu_B)$.

6.4 Numerical illustrations

We turn now to numerical examples to illustrate the behaviour of the two-viscosity fluid. The point here is not to give an accurate specific scheme, which is an interesting perspective since the diffusion term may degenerate, but is beyond the scope of this work. We merely apply here a simple finite volume strategy. The infinite space domain is replaced by some finite computational domain $[a,b]$. Since we do not want to cope with boundary conditions here, we merely impose a free flux on the boundaries, which is compatible with the examples we choose. Positive time and space steps Δt and Δx being given, we introduce the usual notation $t^n = \Delta t$, $n \geqslant 0$, and $x_j = j\Delta x$, $0 \leqslant j \leqslant J$, where $J = (b-a)/\Delta x$. An approximation of the depth h is saught for in the form

$$h_j^{n+1} = h_k^n - \frac{\Delta t}{\Delta x}(F_{j+1/2}^n - F_{j-1/2}^n) + \frac{\Delta t}{\Delta x}(G_{j+1/2}^n - G_{j-1/2}^n),$$

where $F_{j+1/2}^n$ is the numerical advection flux, and $G_{j+1/2}^n$ the numerical diffusion flux, both computed at interface $x_{j+1/2}$. In the following we denote u_j^n the discretized basal velocity, and f_j the discrete topography, which are both given functions.

The advection flux is merely an upwind flux

$$F^n_{j+1/2} = \begin{cases} h^n_j(u^n_j + u^n_{j+1})/2 & \text{if } u^n_j + u^n_{j+1} \geqslant 0, \\ h^n_{j+1}(u^n_j + u^n_{j+1})/2 & \text{if } u^n_j + u^n_{j+1} \leqslant 0. \end{cases}$$

For the diffusive flux, we write $G^n_{j+1/2} = D^n_{j+1/2} K^n_{j+1/2}$, where $K^n_{j+1/2}$ is the approximate value of the slope $\partial_x \varphi$

$$K^n_{j+1/2} = \frac{h^n_{j+1} - h^n_j}{\Delta x} + \frac{f_{j+1} - f_j}{\Delta x},$$

and $D^n_{j+1/2}$ is a discretization of (6.3.15). To obtain it we need to compute h^* and y^* at the interface. Accordingly to (6.3.16), we put

$$(h^*)^n_{j+1/2} = \begin{cases} \dfrac{B}{|K^n_{j+1/2}|} & \text{if } B < \dfrac{h^n_{j+1} + h^n_j}{2} |K^n_{j+1/2}|, \\ \dfrac{h^n_{j+1} + h^n_j}{2} & \text{if not,} \end{cases}$$

and $(y^*)^n_{j+1/2} = (h^n_{j+1} + h^n_j)/2 - ((h^*)^n_{j+1/2}$, so that $D^n_{j+1/2}$ is given by

$$D^n_{j+1/2} = \frac{1}{3}\Bigg(\big((y^*)^n_{j+1/2}\big)^3 + \frac{3}{2}\big(\frac{1}{\nu_B} + 1\big)\big((y^*)^n_{j+1/2}\big)^2 (h^*)^n_{j+1/2}$$

$$+ \frac{3}{\nu_B}(y^*)^n_{j+1/2})\big((h^*)^n_{j+1/2}\big)^2 + \big((h^*)^n_{j+1/2}\big)^3 \Bigg).$$

The time step Δt is actually updated at each time step using the CFL condition

$$\frac{\Delta t^n}{\Delta x^2} = \frac{\sigma}{2D^n}, \qquad \text{with } D^n = \max_j D^n_{j+1/2}, \quad \text{where } \sigma < 1.$$

The following simulations have been performed with $J = 200$ cells in the interval $[-1, 1]$, together with $\sigma = 0.9$. All figures are gathered at the end of the paper.

The first set of simulations concerns the collapse of a square-shaped stack on a horizontal flat bottom: $h^0(x) = 1$ for $x \in]-1/3, 1/3[$, 0 elsewhere, with zero basal velocity ($u_b = v_b = 0$). We first propose a comparison between the two viscosities model and the high viscosity and low viscosity models. The small deformation viscosity is $\nu_B = 100$ (recall that $\nu = 1$), and the yield stress is 0.1 in Figure 6.4, and 0.5 in Figure 6.5. These figures are complemented by Figure 6.6 where we display for four values of the yield stress B a time-lapse of the evolution of both the total thickness of the fluid h (plain lines) and the thickness of the low velocity layer (dashed lines).

For $B = 0.1$, the fluid clearly behaves similarly as the low viscosity fluid in the early stages, then eventually it slows down, when the low viscosity layer tends to disappear, see Figure 6.6, top left. With a yield stress $B = 0.5$, the two viscosities model stays in-between the other two, as expected, faster that the high viscosity model, slower than the low viscosity one, see Figure 6.5. However, one can check that the front hardly moves between $t = 10$ and $t = 50$, indicating that the fluid tends to behave as the high velocity one. This is made more explicit in Figure 6.6, top right, where for $t = 10$ and $t = 50$ the low viscosity layer is very small. In general, the thickness y^* decreases with time, faster when B is larger. It is hardly observable fot $t = 50$ when $B = 2.5$, indicating that the fluid is almost completely driven by the high viscosity.

Using the same initial data, we check the convergence of the two viscosities model towards the Bingham fluid when ν_B goes to ∞. We take a yield stress $B = 1.25$, and $\nu_B = 10, 100, 1000$. As expected, the behaviour becomes close to the Bingham fluid, yet it departs from it for larger times, see Figure 6.7.

We turn now to a different context, closer to the situation in geophysics. The flow here is no longer purely gravity driven, it is actually dragged along by a non zero basal velocity. The idea here is that our pseudo-plastic fluid is a very crude model of some planetary lithosphere, below which lies the mantle. The basal velocity is the upper trace of convection currents in the mantle, which are supposed to be the main drivers of plate tectonics. The initial thickness is constant equal to 1, and we use two basal velocities $U_b(x) = (u_b(x), 0)$, where

$$u_b(x) = -\sin(2\pi x)/10 \cdot \mathbb{1}_{]-0.5,0.5[}(x), \qquad u_b(x) = \sin(2\pi x)/10 \cdot \mathbb{1}_{]-0.5,0.5[}(x). \tag{6.4.1}$$

These velocities crudely correspond respectively to the vertical motion of a magma bubble, which generates local perturbations of the velocity. The first one corresponds to some bubble lift, with negative velocity on the left and positive on the right. It generates some kind of a valley surrounded by mountains, see Figure 6.8. Conversely, the descent of a bubble reverses the velocities, and produces a mountain surrounded with valleys, Figure 6.9. We notice in both cases that the small velocity model has very little influence on the time evolution, and that the two viscosity model leads to rather sharp angles in the thickness.

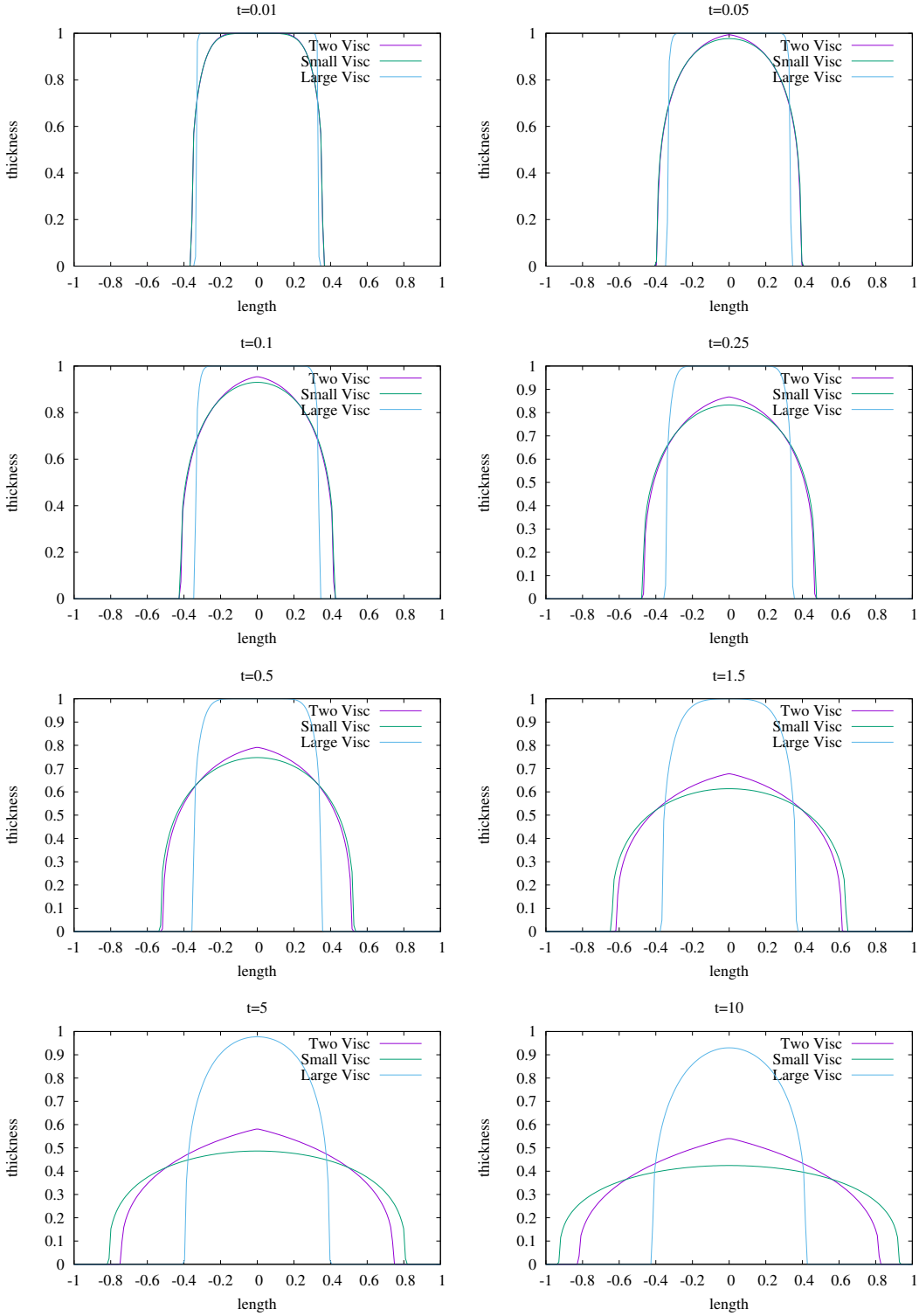


Figure 6.4: Comparison between the three models, time evolution. Yield stress $B = 0.1$, $\nu_B = 100$

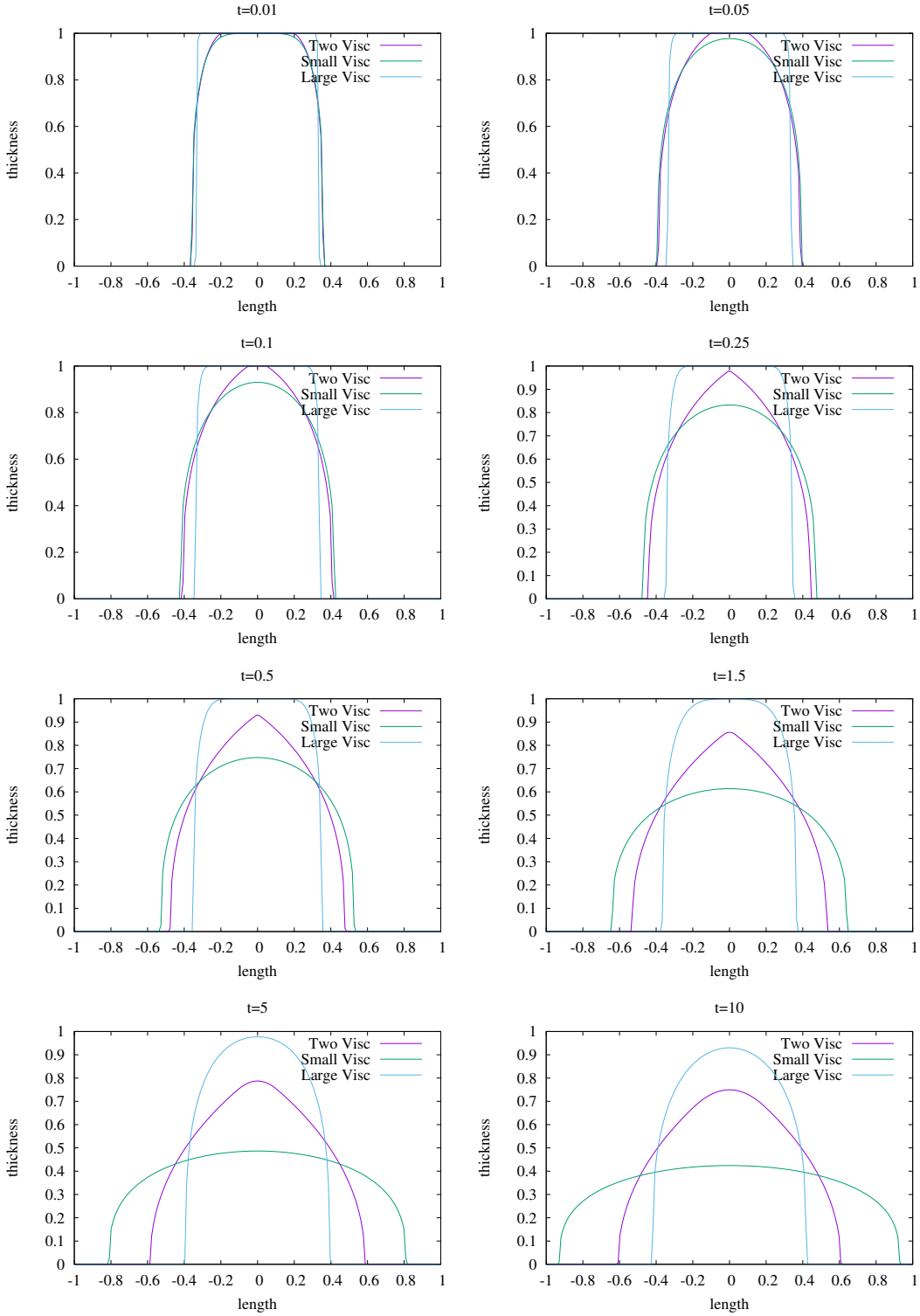


Figure 6.5: Comparison between the three models, time evolution. Yield stress $B = 0.5$, $\nu_B = 100$

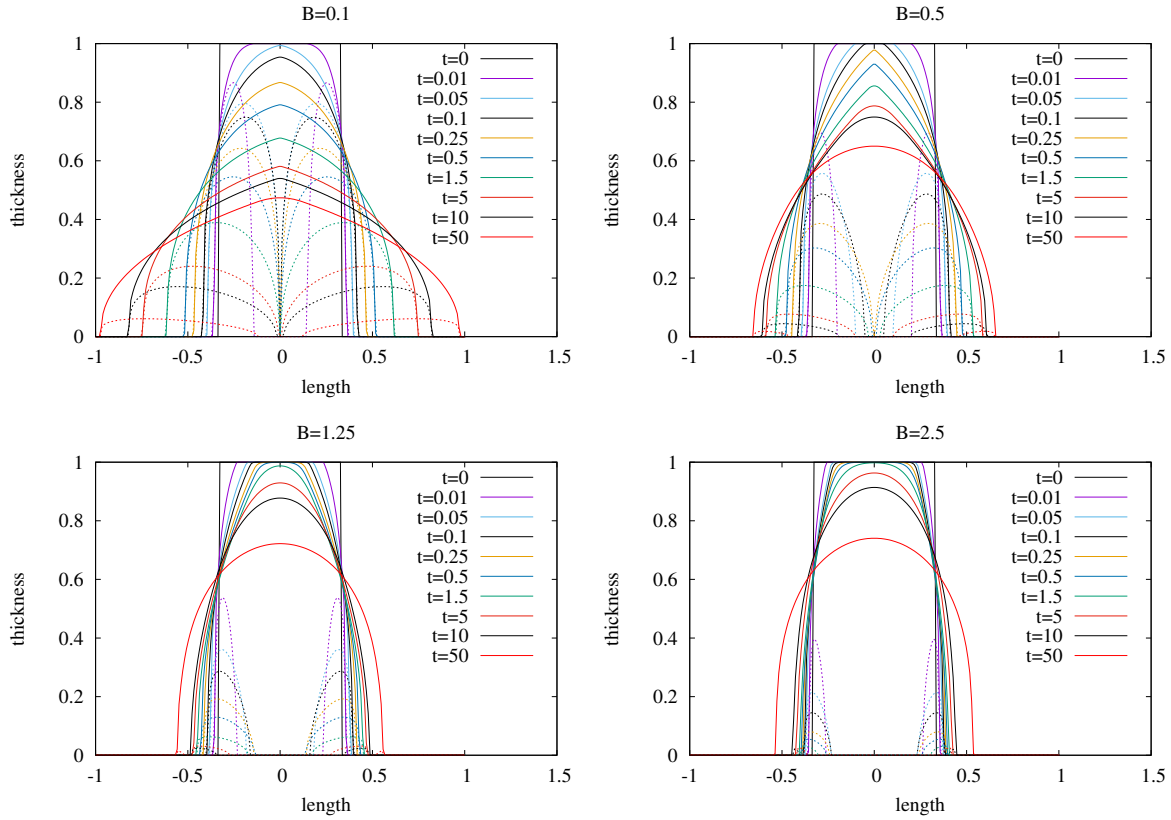


Figure 6.6: Influence of the yield stress B. $\nu_B = 100$. Color code in pictures - plain lines: total thickness h, dashed lines: small viscosity zone thickness y^*.

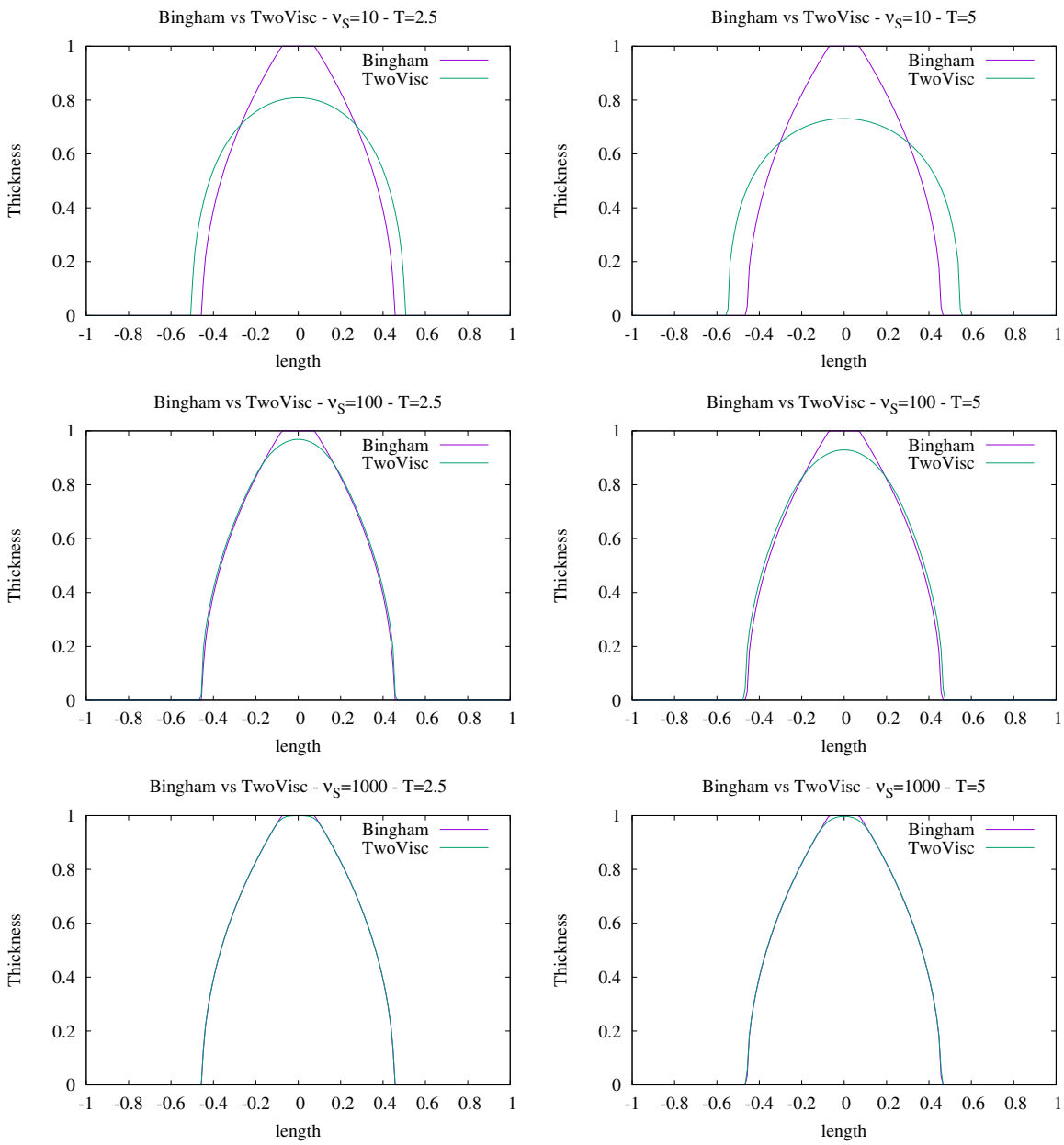


Figure 6.7: Convergence towards the Bingham model. Yield stress $B = 1.25$. From top to bottom: $\nu_B = 10, 100, 1000$.

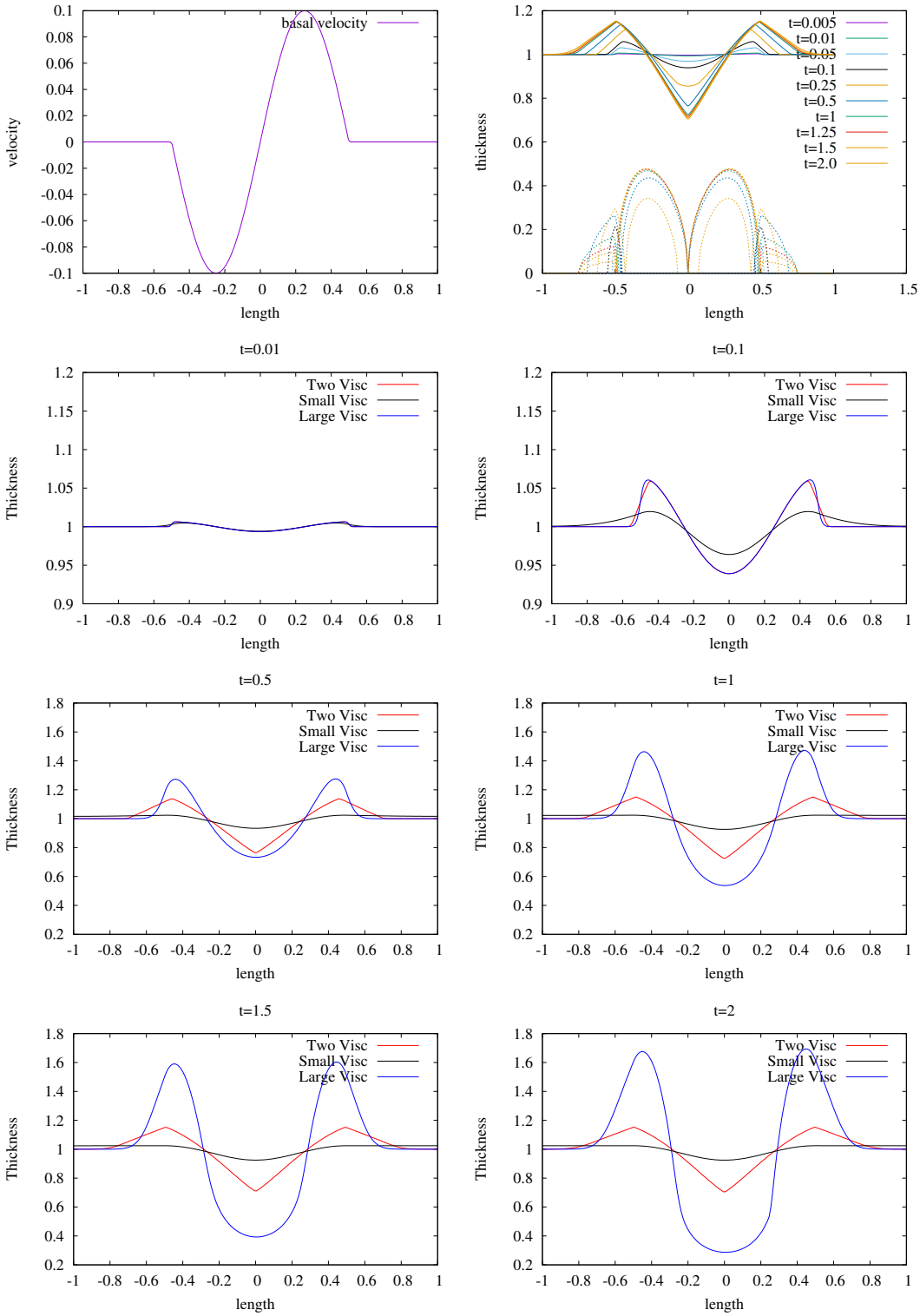


Figure 6.8: Lift of a magma bubble - Yield stress $B = 1.25$ - $\nu_B = 100$ -Top left: basal velocity - Top right: Timelapse oh thickness h and low viscosity layer y^* - Next 6 pictures: time evolution of the three models

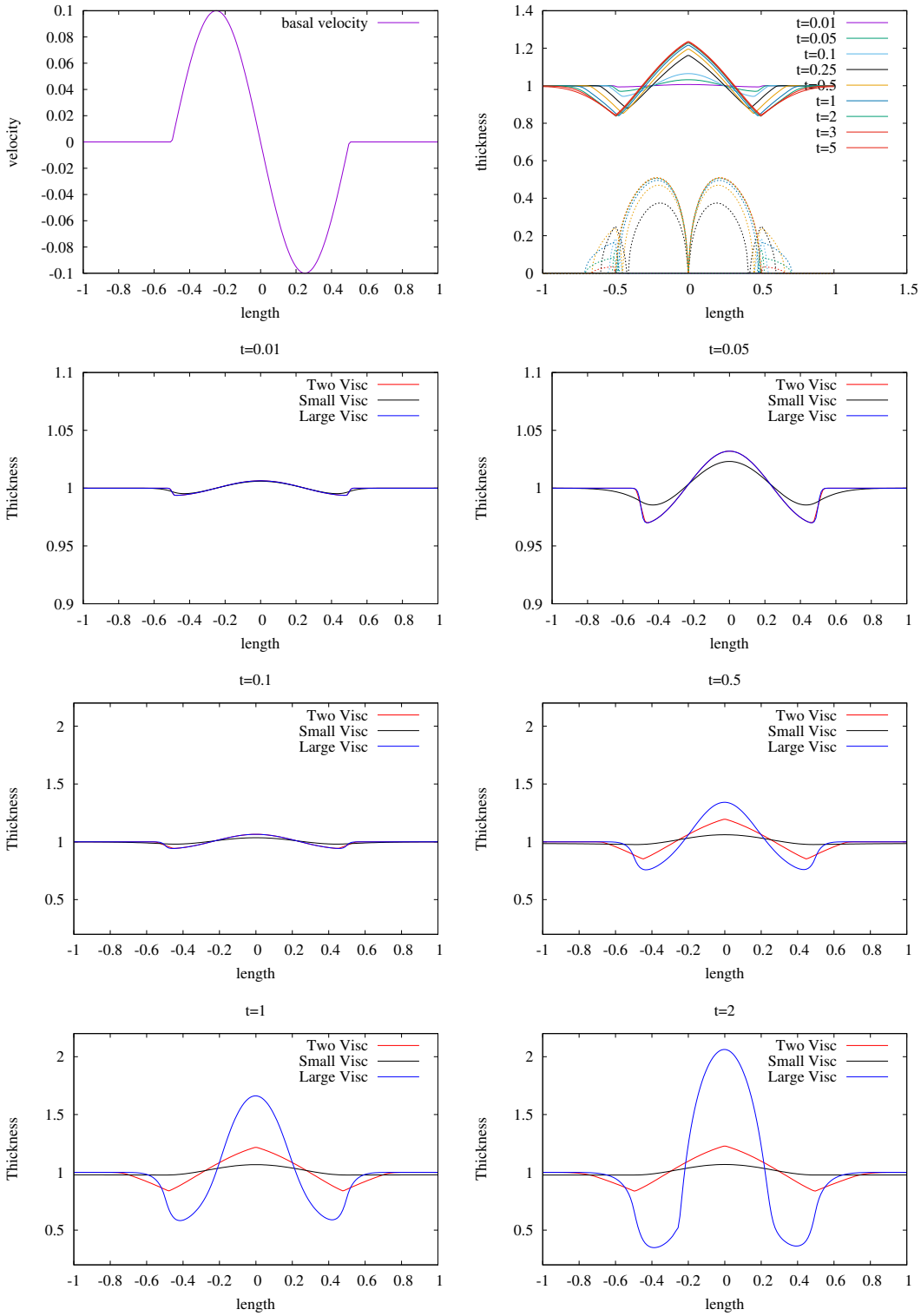


Figure 6.9: Descent of a magma bubble - Yield stress $B = 1.25$ - $\nu_B = 100$ - Top left: basal velocity - Top right: Time-lapse of thickness h and low viscosity layer y^* - Next 6 pictures: time evolution of the three models

Bibliography

[1] Christophe Ancey. Plasticity and geophysical flows: a review. *Journal of Non-Newtonian Fluid Mechanics*, 142(1-3):4–35, 2007.

[2] Neil J Balmforth, Richard V Craster, Alison C Rust, and Roberto Sassi. Viscoplastic flow over an inclined surface. *Journal of Non-Newtonian Fluid Mechanics*, 139(1-2):103–127, 2006.

[3] T Barker, DG Schaeffer, P Bohorquez, and JMNT Gray. Well-posed and ill-posed behaviour of the $\mu(I)-$rheology for granular flow. *Journal of Fluid Mechanics*, 779:794–818, 2015.

[4] T Barker, DG Schaeffer, M Schearer, and JMNT Gray. Well-posed continuum equations for granular flow with compressibility and $\mu(I)-$rheology. *Proceedings of the Royal Society A*, 473:20160846, 2017.

[5] IV Basov and VV Shelukhin. Generalized solutions to the equations of compressible Bingham flows. *Z. Angew. Math. Mech.*, 79:185–192, 1999.

[6] D Bercovici, P Tackley, and Y Ricard. 7.07-the generation of plate tectonics from mantle dynamics. *Treatise on Geophysics. Elsevier, Oxford*, pages 271–318, 2015.

[7] Michel Bercovier and Michael Engelman. A finite-element method for incompressible non-Newtonian flows. *Journal of Computational Physics*, 36(3):313–326, 1980.

[8] Eugene Cook Bingham. An investigation of the laws of plastic flow. US bureau of standards bulletin, 13: 309-353, 1916.

[9] Eugene Cook Bingham. *Fluidity and plasticity*, volume 2. McGraw-Hill, 1922.

[10] David V Boger. Demonstration of upper and lower Newtonian fluid behaviour in a pseudoplastic fluid. *Nature*, 265(5590):126–128, 1977.

[11] François Bouchut. *Nonlinear stability of finite volume methods for hyperbolic conservation laws, and well-balanced schemes for sources.* Springer, 2004.

[12] François Bouchut, David Doyen, and Robert Eymard. Convection and total variation flow. *IMA Journal of Numerical Analysis*, 34(3):1037–1071, 2014.

[13] François Bouchut, David Doyen, and Robert Eymard. Convection and total variation flow-erratum and improvement. *IMA Journal of Numerical Analysis*, 37(4):2139–2169, 2017.

[14] François Bouchut, Robert Eymard, and Alain Prignet. Convergence of conforming approximations for inviscid incompressible Bingham fluid flows and related problems. *Journal of Evolution Equations*, 14(3):635–669, 2014.

[15] Didier Bresch, Enrique D Fernandez-Nieto, Ioan R Ionescu, and Paul Vigneaux. Augmented Lagrangian method and compressible visco-plastic flows: applications to shallow dense avalanches. In *New directions in mathematical fluid mechanics*, pages 57–89. Springer, 2009.

[16] Haim Brezis. *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, volume 5. Elsevier, 1973.

[17] C Carstensen, F Ebobisse, AT McBride, BD Reddy, and P Steinmann. Some properties of the dissipative model of strain-gradient plasticity. *Philosophical Magazine*, 97(10):693–717, 2017.

[18] C Carstensen, BD Reddy, and M Schedensack. A natural nonconforming FEM for the Bingham flow problem is quasi-optimal. *Numerische Mathematik*, 133(1):37–66, 2016.

[19] Renaud Chalayer, Laurent Chupin, and Thierry Dubois. A bi-projection method for incompressible Bingham flows with variable density, viscosity, and yield stress. *SIAM Journal on Numerical Analysis*, 56(4):2461–2483, 2018.

[20] Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of mathematical imaging and vision*, 40(1):120–145, 2011.

[21] Maria Chatzimina, Georgios C Georgiou, Ioannis Argyropaidas, Evan Mitsoulis, and RR Huilgol. Cessation of Couette and Poiseuille flows of a Bingham plastic and finite stopping times. *Journal of non-Newtonian fluid mechanics*, 129(3):117–127, 2005.

[22] Laurent Chupin and Thierry Dubois. A bi-projection method for Bingham type flows. *Computers & Mathematics with Applications*, 72(5):1263–1286, 2016.

[23] Laurent Chupin and Jordane Mathe. Existence theorem for homogeneous incompressible Navier-Stokes equation with variable rheology. *European Journal of Mechanics B/Fluids*, 61:135–143, 2017.

[24] Richard Courant, Kurt Friedrichs, and Hans Lewy. On the partial difference equations of mathematical physics. *IBM journal of Research and Development*, 11(2):215–234, 1967.

[25] Georges Duvaut and Jacques-Louis Lions. Les inéquations en mécanique et en physique. dunod, paris, 1972. *Travaux et Recherches Mathématiques*, (21), 1976.

[26] Georges Duvaut and Jacques-Louis Lions. *Inequalities in mechanics and physics*, volume 219. Springer Science & Business Media, 2012.

[27] Eduard Feireisl, X Liao, and Josef Malek. Global weak solutions to a class of non-Newtonian compressible fluids. *Mathematical Methods in the Applied Sciences*, 38(16):3482–3494, 2015.

[28] Michel Fortin and Roland Glowinski. Augmented Lagrangian methods in quadratic programming. In *Studies in Mathematics and its Applications*, volume 15, pages 1–46. Elsevier, 1983.

[29] Roland Glowinski, Jinchao Xu, and Philippe G. Ciarlet, editors. *Handbook of numerical analysis. Vol. XVI. Special volume: Numerical methods for non-Newtonian fluids*, volume 16 of *Handbook of Numerical Analysis*. Elsevier/North-Holland, Amsterdam, 2011.

[30] Tom Goldstein, Min Li, Xiaoming Yuan, Ernie Esser, and Richard Baraniuk. Adaptive primal-dual hybrid gradient methods for saddle-point problems. *arXiv preprint arXiv:1305.0546*, 2013.

[31] Jean-Luc Guermond, Peter Minev, and Jie Shen. An overview of projection methods for incompressible flows. *Computer methods in applied mechanics and engineering*, 195(44-47):6011–6045, 2006.

[32] Weimin Han and B Daya Reddy. Numerical analysis of the primal problem. In *Plasticity*, pages 319–370. Springer, 2013.

[33] JW He and Roland Glowinski. Steady Bingham fluid flow in cylindrical pipes: a time dependent approach to the iterative solution. *Numerical linear algebra with applications*, 7(6):381–428, 2000.

[34] WH Herschel and Ronald Bulkley. Measurement of consistency as applied to rubber-benzene solutions. In *Am. Soc. Test Proc*, volume 26, pages 621–633, 1926.

[35] Herbert E Huppert. The propagation of two-dimensional and axisymmetric viscous gravity currents over a rigid horizontal surface. *Journal of Fluid Mechanics*, 121:43–58, 1982.

[36] François James, Pierre-Yves Lagrée, Minh H Le, and Mathilde Legrand. Towards a new friction model for shallow water equations through an interactive viscous layer. *ESAIM: Mathematical Modelling and Numerical Analysis*, 53(1):269–299, 2019.

[37] Pierre Jop, Yoël Forterre, and Olivier Pouliquen. A constitutive law for dense granular flows. *Nature*, 441(7094):727–730, 2006.

[38] Pierre-Yves Lagrée, Lydie Staron, and Stéphane Popinet. The granular column collapse as a continuum: validity of a two-dimensional Navier–Stokes model with a $\mu(I)-$ rheology. *Journal of Fluid Mechanics*, 686:378–408, 2011.

[39] Jacques-Louis Lions. Remarks on some nonlinear evolution problems arising in Bingham flows. *Israel Journal of Mathematics*, 13(1-2):155–172, 1972.

[40] KF Liu and CC Mei. Approximate equations for the slow spreading of a thin sheet of Bingham plastic fluid. *Physics of Fluids A: Fluid Dynamics*, 2(1):30–36, 1990.

[41] Ko Fei Liu and Chiang C Mei. Slow spreading of a sheet of Bingham fluid on an inclined plane. *Journal of fluid mechanics*, 207:505–529, 1989.

[42] Christelle Lusso, François Bouchut, Alexandre Ern, and Anne Mangeney. Explicit solutions to a free interface model for the static/flowing transition in thin granular flows. *ESAIM: Math. Modelling Numer. Anal.*, 2020.

[43] Christelle Lusso, Alexandre Ern, François Bouchut, Anne Mangeney, Maxime Farin, and Olivier Roche. Two-dimensional simulation by regularization of free surface viscoplastic flows with Drucker–Prager yield stress and application to granular collapse. *Journal of Computational Physics*, 333:387–408, 2017.

[44] Josef Málek and KR Rajagopal. Compressible generalized Newtonian fluids. *Zeitschrift für angewandte Mathematik und Physik*, 61(6):1097–1110, 2010.

[45] AE Mamontov. Existence of global solutions to multidimensional equations for Bingham fluids. *Mathematical Notes*, 82(4):501–517, 2007.

[46] Evan Mitsoulis and Th Zisis. Flow of Bingham plastics in a lid-driven square cavity. *Journal of non-newtonian fluid mechanics*, 101(1-3):173–180, 2001.

[47] Yurii Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer Science & Business Media, 2013.

[48] Osborne Reynolds. IV. On the theory of lubrication and its application to Mr. Beauchamp Tower's experiments, including an experimental determination of the viscosity of olive oil. *Philosophical transactions of the Royal Society of London*, (177):157–234, 1886.

[49] Ralph Tyrell Rockafellar. *Convex analysis*. Princeton university press, 2015.

[50] Nicolas Roquet and Pierre Saramito. An adaptive finite element method for Bingham fluid flows around a cylinder. *Computational Methods in Applied Mechanics and Engineering*, 192:3317–3341, 2003.

[51] S. Roux and F. Radjai. Texture-dependent rigid plastic behaviour. In H. J. Herrmann et al., editor, *Physics of Dry Granular Media*, volume 350 of *NATO ASI Series*, pages 229–236. Springer, 1998.

[52] Oxana Sadovskaya and Vladimir Sadovskii. *Mathematical modeling in mechanics of granular materials*, volume 21. Springer Science & Business Media, 2012.

[53] Pierre Saramito. Méthodes numériques en fluides complexes: théorie et algorithmes. 2013.

[54] Pierre Saramito. A damped Newton algorithm for computing viscoplastic fluid flows. *Journal of Non-Newtonian fluid mechanics*, 238:6–15, 2016.

[55] Pierre Saramito and Anthony Wachs. Progress in numerical simulation of yield stress fluid flows. *Rheologica Acta*, 56(3):211–230, 2017.

[56] Paul J Tackley. Self-consistent generation of tectonic plates in time-dependent, three-dimensional mantle convection simulations 2. Strain weakening and asthenosphere. *Geochemistry, Geophysics, Geosystems*, 1(8), 2000.

[57] Roger Temam. *Problèmes mathématiques en plasticité*. Gauthier-Villars, Montrouge, 1983.

[58] Raymond Trémolières, Jacques-Louis Lions, and Roland Glowinski. *Numerical analysis of variational inequalities*. Elsevier, 2011.

[59] Milton Van Dyke. Perturbation methods in fluid mechanics/annotated edition. *NASA STI/Recon Technical Report A*, 75, 1975.

[60] Shijie Zhong, Michael Gurnis, and Louis Moresi. Role of faults, nonlinear rheology, and viscosity structure in generating plates from instantaneous mantle flow models. *Journal of Geophysical Research: Solid Earth*, 103(B7):15255–15268, 1998.